

# PREVISÃO MULTI-PASSOS EM SÉRIES TEMPORAIS: ESTRATÉGIAS CLÁSSICAS E DE APRENDIZAGEM AUTOMÁTICA

por

Geslie Fernandes Costa

Tese de Mestrado em Análise de Dados e Sistemas de Apoio à Decisão

Orientada por

Prof. Dra Maria Eduarda da Rocha Pinto Augusto da Silva

Co Orientada por

Prof. Dr João Manuel Portela da Gama

**Faculdade de Economia**

Universidade do Porto

2015

Dedico a Deus, razão da minha existência, por fazer novas todas às coisas e pelas  
maravilhas que faz na minha vida.  
Ao meu amor, companheiro, amigo e marido, Juvenal.  
Aos meus pais por me apoiar sempre.

# Nota Biográfica

Geslie Fernandes Costa nasceu a 02 de Setembro de 1984, fez a sua formação básica e secundária no estado de São Paulo, Brasil. Licenciou-se em Tecnologia e Sistemas de Informação pela Faculdade de Estácio UniRadial, iniciando o Mestrado em Análise de Dados e Sistemas de Apoio à Decisão no ano de 2013.

Profissional com cerca de 9 anos de experiência atuando nas áreas análise e desenvolvimento de sistemas em empresas de diversos setores de atividades. Atualmente é Consultora da Plataforma de Dados da Saúde (PDS) na Altran Portugal.

# Agradecimentos

A todos que colaboraram e apoiaram na realização deste trabalho, dentre os quais desejo agradecer em especial.

Aos meus pais queridos, Raimundo e Cleidimar que são minha fonte de inspiração e cuidado, pela ajuda que me deram ao longo desses anos, fazendo-me uma pessoa melhor.

Aos meus amados irmãos, Washington e William, pelo apoio constante e incentivo ao estudo. Vocês são partes fundamentais na minha formação como pessoa e profissional.

Á minha irmã Gisele, que Deus colocou no meu caminho para partilhar meus risos e enxugar minhas lágrimas.

Ao meu amor e amigo, Juvenal, que me ajudou em todos os momentos, principalmente nos dias de desânimo e cansaço, ao longo desse tempo demonstrando muito carinho e compreensão.

A minha orientadora e professora Dra Maria Eduarda da Rocha Pinto Augusto da Silva, pela excelente orientação do trabalho, pela compreensão e paciência durante todo esse tempo. Obrigada por sempre ouvir meus questionamentos e dúvidas.

Ao meu co orientador e professor Dr João Manuel Portela da Gama, pela orientação competente e pela disposição em servir com seu melhor.

Aos amigos do MADSAD, Rui Filipe e Sônia Teixeira, obrigada pela companhia e amizade durante os dois anos de mestrado, pelas dicas, esclarecimentos e pelas valiosas contribuições no meu trabalho.

# Tabela de Notação

Notação	Descrição
$Y_T = y_1, \dots, y_T$	Série temporal com $T$ observações
$H$	Número dos horizontes de previsão requeridos
$h$	Horizonte da previsão
$p$	Ordem do modelo estimado
$\beta$	Vetor de parâmetros
$\Psi$	Vetor de hiperparâmetros
$m^{(h)}$	Modelo $m$ utilizado em $h$ tempo de forma recursiva
$m_h$	Modelo direto para o horizonte $h$
$\mathbf{x}_t = [y_t, \dots, y_{t-d}]'$	Vetor de entradas no tempo $t$ com $d$ variáveis desfasadas
$\mathbf{z}_t = [y_t, \dots, y_{t-p}]'$	Vetor de entradas no tempo $t$ com $p$ variáveis desfasadas
$\mathbf{r}_t = [y_t, \dots, y_{t-p_h}]'$	Vetor de entradas no tempo $t$ com $p_h$ variáveis desfasadas
$\phi$	Vetor de parâmetros para o modelo um-passo à frente
$\gamma$	Vetor de parâmetros para o modelo direto
REC	Estratégia definida na equação (2.21)
DIR	Estratégia definida na equação (2.25)
RTY	Estratégia definida na equação (2.28)
LIN	Modelo Linear
MLP	Modelo Rede Neuronal
BST-L2	Modelo Gradiente boosting com função de perda $L_2$
BST-L1	Modelo Gradiente boosting com função de perda $L_1$
AR	Processo linear autoregressivo

# Resumo

A previsão dos valores futuros de uma série temporal, designada por previsões multi-passos, reveste-se de muita importância uma vez que permite o planeamento antecipado e correto dos sistemas em estudo. A previsão em séries temporais tem sido influenciada principalmente pela econometria e estatística. Nas últimas décadas, a aprendizagem automática, do inglês *machine learning*, uma área dedicada ao desenvolvimento de algoritmos que *aprendem* a partir dos dados, tem-se mostrado muito ativa em modelação e predição. No entanto, tem havido pouca investigação que relacione aprendizagem automática e estratégias de previsão multi-passos em séries temporais. O objetivo desta dissertação é contribuir para colmatar essa lacuna, abordando o problema de previsão multi-passos em séries temporais a partir da perspectiva das estratégias clássica e de aprendizagem automática. A literatura apresenta duas principais abordagens à previsão multi-passos: a recursiva e direta. Na abordagem recursiva, estima-se um único modelo para a série temporal que produz previsões um-passo à frente recursivamente. Na abordagem direta, especifica-se e estima-se um modelo diferente para cada horizonte de previsão. Uma outra abordagem introduzida recentemente constitui uma forte candidata a solucionar problemas da previsão multi-passo. Esta estratégia, denominada retificação, combina as duas abordagens clássicas ajustando o resultado de uma previsão recursiva linear com um modelo de retificação não linear usando o algoritmo de gradiente *boosting*. Para além de evitar a escolha entre as previsões recursivas e diretas, em alguns casos os resultados desta abordagem demonstram ser melhor, ou pelo menos tem um desempenho próximo, a recursivas e diretas. Nesta dissertação compararam-se as abordagens recursivas, diretas e de retificação, com modelos gerados por métodos estatísticos e de aprendizagem automática, utilizando séries temporais reais com diferentes comportamentos e horizontes de previsão.

**Palavras-Chave:** séries temporais, previsão multi-passos, aprendizagem automática, rede neuronal, gradiente *boosting*, ARIMA, GARCH

# Abstract

Forecasting future observations from a time series is a problem of increasing importance. Time series forecasting has been influenced mainly by econometrics and statistics. In the last decades, machine learning, an area dedicated to the development of algorithms that learn from data, has become one of the most active areas of predictive modeling research. However, there has been little research relating machine learning and time series forecasting. The goal of this dissertation is to approach forecasting problems from both perspectives: classic strategies and machine learning. The literature presents two main approaches to multi-step forecasts: recursive and direct. In the recursive approach, a single model for the observed time series is estimated. This model produces optimal one-step-ahead predictions which are iterated to produce multi-step-ahead forecasts. The direct approach specifies a different model for each required forecasting horizon. Recently, another approach, combining the direct and recursive strategies, has been proposed in the literature: the rectify algorithm. Briefly, this algorithm starts with linear recursive forecasts which are then adjusted (rectified) with a nonlinear model using a gradient boosting algorithm. Therefore, this approach avoids the choice between recursive and direct forecasts and presents a performance that is not worse than the classical ones. This dissertation compares the performance of recursive, direct and rectify approaches in forecasting observed time series with different characteristics and forecast horizons. **Keywords:** time series, multi-step forecast, machine learning, neural network, gradient boosting, ARIMA, GARCH.

# Índice

Nota Biográfica	ii
Agradecimentos	iii
Tabela de Notação	iv
Resumo	v
Abstract	vi
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	2
1.2 Objetivos . . . . .	3
1.3 Organização . . . . .	3
<b>2 Revisão da Literatura</b>	<b>4</b>
2.1 Séries Temporais . . . . .	4
2.2 Modelos . . . . .	7
2.2.1 Modelos Estatísticos . . . . .	9
2.2.2 Aprendizagem Automática . . . . .	11
2.3 Seleção de Modelos . . . . .	15
2.4 Métricas de Avaliação . . . . .	17
2.5 Previsão Multi-Passos . . . . .	18
2.5.1 Estratégia Recursiva . . . . .	19
2.5.2 Estratégia Direta . . . . .	20
2.5.3 Estratégia de Retificação . . . . .	20
<b>3 Casos de Estudo</b>	<b>23</b>
3.1 Caso 1 - Consumo de Energia Elétrica . . . . .	25
3.2 Caso 2 - Retornos do New York Stock Exchange . . . . .	30
3.3 Caso 3 - Índice de Produção Mensal do <i>Federal Reserve Board</i> . . . . .	34
3.4 Considerações Finais . . . . .	40



<b>4</b>	<b>Conclusões</b>	<b>41</b>
4.1	Limitações e Trabalho Futuro . . . . .	42
	<b>Bibliografia</b>	<b>43</b>
	<b>Anexo</b>	<b>48</b>
<b>A</b>	<b>Estratégias de Previsão Multi-passos</b>	<b>48</b>
A.1	Estratégia Recursiva . . . . .	48
A.2	Estratégia Direta . . . . .	49
<b>B</b>	<b>Estudo de Casos</b>	<b>51</b>
B.1	Resultados das Previsões Multi-Passos . . . . .	51

## Lista de Tabelas

B.1	Valor real observado com os valores previstos do consumo de energia elétrica ao longo dos horizontes de previsões vinte quatro passos à frente $H = 24$ em Cabinda . . . . .	52
B.2	Medidas de precisão ao longo de diferentes horizontes para previsões vinte e quatro passos à frente no consumo de energia elétrica em Cabinda. . . . .	53
B.3	Medidas de precisão globais das previsões para as diferentes estratégias no consumo de energia elétrica em Cabinda. . . . .	54
B.4	Valor real observado juntamente com os valores previstos dos retornos da NYSE ao longo dos horizontes de previsões cinco à frente $H = 5$ . .	54
B.5	Medidas de precisão ao longo dos horizontes de previsões cinco passos à frente $H = 5$ dos retornos da NYSE. . . . .	55
B.6	Medidas de precisão das diferentes estratégias e modelos de previsão para $H = 5$ dos retornos da NYSE. . . . .	55
B.7	Valor real observado juntamente com os valores previstos dos índices de produção mensal do Sistema de Reserva Federal com $H = 12$ . . . .	56
B.8	Medidas de precisão ao longo dos horizontes de previsões doze passos à frente $H = 12$ do índice de produção mensal do Sistema de Reserva Federal. . . . .	56
B.9	Medidas de precisão das diferentes estratégias e modelos de previsão para $H = 12$ do índice de produção mensal. . . . .	57

# Lista de Figuras

2.1	Exemplo de séries temporais (Valores e Retornos) com as cotações das ações de uma empresa em minutos. . . . .	8
2.2	Exemplo ilustrativo de uma rede MLP . . . . .	12
2.3	Exemplo para estimação <i>smooth effect</i> usando apenas a variável $x1$ (esquerda) e a variável $x2$ com classes <i>smooth effect</i> ( $knots = quantile(x2, c(0.25, 0.5, 0.75))$ , $df=5$ ) (direita). Assim, o <i>P-splines</i> da esquerda usa por defeito o valor de 20 <i>knots</i> e 4 graus de liberdade, enquanto o <i>P-splines</i> estimado da direita é derivado de 3 <i>knots</i> colocados nos quartis e 5 graus de liberdade. Fonte Hofner et al. (2014). . . . .	15
2.4	Ilustração da ocorrência de: (a) sobre-ajuste, (b) sub-ajuste, (c) função de aproximação mais adequada. . . . .	15
2.5	Validação cruzada de séries temporais, onde cada conjunto de treino (azul) contem uma observação a mais que o conjunto anterior, e consequentemente cada conjunto de teste (vermelho) tem menos uma observações que o anterior. Fonte (Hyndman, 2014). . . . .	16
2.6	Validação cruzada de séries temporais com base em previsões de um-passo. Os pontos azuis representam os conjuntos de treino, os vermelhos de teste e os pontos cinzentos são ignorados. Fonte (Hyndman, 2014). . . . .	17
3.1	Consumo médio (MWh) horário de eletricidade no período que se estende de janeiro de 2011 à dezembro de 2014. . . . .	25
3.2	Consumo médio (MWh) diário no período de janeiro de 2011 à dezembro de 2014. . . . .	26
3.3	Gráficos de <i>ACF</i> e <i>PACF</i> da série de energia elétrica com os limites do intervalo de confiança em 95% (tracejado em azul). . . . .	26
3.4	Consumo na hora $t$ versus o consumo desfasado em diferentes <i>lags</i> ( $t - i$ com $i = 1, \dots, 48$ ), indica que há uma forte dependência para a mesma hora do dia $t - 24$ (no dia anterior) e $t - 48$ (dois dias antes). . . . .	27
3.5	Gráficos dos resultados das diferentes estratégias para comparação dos valores previstos versus observados (cor preta) do consumo de energia elétrica em 24 passos à frente. . . . .	29

3.6	Gráficos dos índices MAPE(%) que avaliaram o desempenho das previsões nas diferentes estratégias e modelos no consumo de energia elétrica com $H = 24$ . . . . .	30
3.7	Retornos da NYSE. Os dados são valores ponderados dos retornos no mercado, a partir de 02 de fevereiro de 1984 à 31 de dezembro de 1991 (2000 dias de negociação). 19 de outubro de 1987 ocorre em $t = 938$ . . . . .	31
3.8	Gráficos de $ACF$ e $PACF$ dos retornos da NYSE, com os limites do intervalo de confiança em 95% (tracejado em azul). . . . .	31
3.9	Gráficos dos resultados das diferentes estratégias para comparação dos valores previstos versus observados (cor preta) dos retornos da NYSE em 5 passos à frente. . . . .	33
3.10	Gráficos dos índices RMSE que avaliaram o desempenho das previsões nas diferentes estratégias e modelos dos retornos da NYSE com $H = 5$ . . . . .	34
3.11	Índice de produção mensal do Sistema de Reserva Federal (1948-1978, $T = 372$ meses). . . . .	35
3.12	Decomposição dos principais componente (sazonal, tendência e ciclo) do índice de produção mensal do Sistema de Reserva Federal (1948-1978). . . . .	35
3.13	Gráfico da $ACF$ do índice de produção, com os limites do intervalo de confiança em 95% (tracejado em azul) e $lag$ máximo de 48. . . . .	36
3.14	Gráficos de $ACF$ e $PACF$ do índice de produção com uma diferença simples e os limites do intervalo de confiança em 95% (tracejado em azul) e $lag$ máximo de 48. . . . .	37
3.15	Gráficos dos resultados das diferentes estratégias e modelos para comparação dos valores previstos versus observados do índice de produção mensal do Sistema de Reserva Federal 12 passos à frente. . . . .	38
3.16	Gráficos dos índice MAPE (%) que avaliaram o desempenho das previsões nas diferentes estratégias e modelos do índice de produção mensal do Sistema de Reserva Federal com $H = 12$ . . . . .	39

# Capítulo 1

## Introdução

Atualmente, a previsão de valores futuros de uma variável de interesse usando apenas os valores observados no passado, constitui um dos desafios mais importantes da análise de dados em muitos cenários da vida real. Este problema pertence ao domínio da análise de séries temporais. Distingue-se previsão a um-passo quando pretende-se prever apenas a próxima observação e multi-passos quando objetivo é obter previsões para vários momentos no futuro. Segundo [De Gooijer and Hyndman \(2006\)](#), a análise da previsão em séries temporais, iniciou entre as décadas de 1970 e 1980 e foi influenciado ao longo dos anos principalmente pelos métodos estatísticos lineares, como o modelo ARIMA (*Autoregressive Integrated Moving Average*). No entanto, tornou-se cada vez mais claro que os modelos lineares não estão adaptados para muitas aplicações reais. No mesmo período, alguns modelos não lineares foram propostos, como: modelo auto-regressivo limiar, introduzido por [Tong and Lim \(1980\)](#) e o modelo ARCH (*Autoregressive Conditional Heteroscedastic*), criado por [Engle \(1982\)](#). É grande a diferença no número de estudos comparando os modelos lineares e não lineares. [De Gooijer and Hyndman \(2006\)](#), revelam que de modo geral, a previsão de modelos não lineares ainda está na sua infância em comparação com a previsão de modelos lineares, nomeadamente devido à sua complexidade analítica e computacional. Com características específicas, a previsão em séries temporais despertou nas duas últimas décadas, o interesse de pesquisadores na área da extração de conhecimento (KDD, do inglês, *Knowledge Discovery in Databases*), onde começaram a desenvolver métodos computacionais de previsão para séries temporais recorrendo a algoritmos de aprendizagem automática. Esses algoritmos tem-se tornando extremamente úteis para resolver problemas importante em grande escala, com uma vasta influência em nossa sociedade atual e futura ([Gama et al., 2012](#)). No entanto, estudos envolvendo estes modelos não se têm debruçado sobre previsões multi-passos, relata [Kline \(2004\)](#). De facto, poucos estudos aprofundados consideraram algoritmos de aprendizagem automática, considerando, por vezes, apenas um algoritmo de redes neurais ou está limitada a previsão de um-passo à frente. Por exemplo, [Ahmed et al. \(2010\)](#), apresentam uma comparação empírica com mode-

los de aprendizagem automática (redes neurais, *kernel regression*, *k*-vizinhos mais próximos, árvore de regressão, *support vector machine* e processos gaussianos) para previsões em séries temporais. Os autores comparam diferentes modelos aliados a séries temporais da competição M3, no entanto, o estudo foi limitado apenas ao problema de previsão um-passo à frente.

A literatura distingue duas estratégias para previsões multi-passos, nomeadamente a *estratégia recursiva* e a *direta*. A *estratégia recursiva*, também conhecida como iterada é considerada a mais antiga e utiliza a recursividade para calcular as previsões. Diferente da *estratégia recursiva*, a *direta* não utiliza a recursão e sim diferentes modelos são gerados para cada horizonte de previsão, mas sempre tendo como entrada os dados reais, (Weigend, 1994; Sorjamaa et al., 2007; Cheng et al., 2006; Tiao and Tsay, 1994; Kline, 2004; Hamzaçebi et al., 2009). Recentemente Taieb and Hyndman (2014), elaboraram a estratégia de retificação com o algoritmo gradiente *boosting* com a função de perda *L2*, esta tem como principal ideia combinar as melhores propriedades da *estratégia recursiva* e da *direta*. Considerando cerca de 500 series temporais das competições de previsão M3 e NN5 (com diferentes comportamentos, tamanhos e horizontes), o estudo conclui que esta estratégia obteve bons resultados.

Desenvolver e comparar diferentes estratégias de previsões multi-passos baseados em modelos lineares e não lineares é uma tarefa difícil, uma vez que fatores como o comprimento da série, o horizonte da previsão e o processo para geração dos dados são fundamentais para os modelos produzirem boas previsões.

## 1.1 Motivação

A obtenção de previsões multi-passos é um tema que desempenha um papel importante em diversas áreas da atividade científica, industrial e económica, como em meteorologia, telecomunicações e finanças (Palit and Popovic, 2006). A literatura sobre análise de séries temporais apresenta diversas ferramentas para a modelação e previsão deste tipo de dados. No entanto, observa-se na literatura especializada, uma crescente necessidade de estudos sobre outros tipos de estratégias que possam ser aplicadas em previsão a longo prazo de séries temporais (Ben Taieb et al., 2012). Extrair conhecimento a partir dos dados é um objetivo comum entre as áreas da estatística e aprendizagem automática. A aprendizagem automática é um campo relativamente jovem que tem evoluído na comunidade de ciência da computação. Já a estatística é um campo mais antigo que tem muito em comum com a aprendizagem automática, mas evoluiu no campo da matemática. Unir estas duas abordagens na previsão multi-passos em séries temporais tem sido um problema em aberto na literatura especializada.

## 1.2 Objetivos

O principal objetivo desta dissertação será comparar diferentes estratégias de previsões multi-passos à frente com modelos lineares e não lineares, gerados por métodos estatísticos e algoritmos de aprendizagem automática, nomeadamente, ARIMA, GARCH, rede neuronal e gradiente *boosting*. Tradicionalmente, as estratégias recursivas e diretas são as mais utilizadas em séries temporais, geralmente do um ponto de vista concorrencial. Por outro lado, [Taieb and Hyndman \(2014\)](#) introduziram uma estratégia de retificação que concilia as duas estratégias anteriores gerando previsões recursivas com um modelo linear, ajustando-as, posteriormente, com um modelo direto não linear. Neste trabalho, reproduz-se essa estratégia de retificação e aplica-se em três casos de estudo, constituídos por séries temporais com características variadas. O primeiro caso refere-se ao consumo horário de energia elétrica, uma série com sazonalidades múltiplas, 35064 observações e o objetivo será prever o horizonte de 24 passos à frente. O segundo, apresenta uma série com dados do mercado financeiro, 2000 observações diárias e um horizonte de previsão 5 passos à frente. O terceiro e último caso trata de uma série temporal de índices macroeconómicos, esta apresenta tendência linear, sazonalidade periódica e o objetivo será prever 12 passos à frente. São obtidas e avaliadas previsões produzidas pelas estratégias recursivas, diretas e de retificação com modelos e lineares não lineares.

## 1.3 Organização

Este trabalho encontra-se dividido em quatro capítulos, seguindo uma sequência lógica para o desenvolvimento do tema abordado. O capítulo 2, consiste numa breve descrição dos conceitos básicos e essenciais relacionados com tema: Séries Temporais, Modelos, Seleção de Modelos, Métricas de Avaliação e Estratégias de Previsão Multi-Passos. No capítulo 3 são apresentados três casos de estudo, onde foram geradas as previsões multi-passos em séries temporais reais com três estratégias diferentes. Foram ainda utilizadas três métricas para validação dos resultados. A linguagem R ([R Core Team, 2014](#)) foi a escolhida para elaboração de todo o terceiro capítulo. Por fim, no capítulo 4 são apresentadas as principais conclusões e sugestões de trabalho futuro.

# Capítulo 2

## Revisão da Literatura

A utilização de métodos e técnicas de análise de séries temporais para representar informações intrínsecas dessas séries, possibilita agrupar, classificar, compreender e prever eventos futuros. Dentro deste contexto, neste capítulo, serão brevemente descritos os conceitos necessários a este trabalho. Assim na Secção 2.1 introduzem-se definições fundamentais de séries temporais, na Secção 2.2, aborda-se modelação de dados com modelos estatísticos e de aprendizagem automática, na Secção 2.3 discutem-se procedimentos para seleção de modelos, na Secção 2.4 consideram-se métricas para avaliação de previsões e por fim na Secção 2.5, descrevem-se diversas estratégias para obtenção de previsões multi-passos.

### 2.1 Séries Temporais

Uma série temporal é um conjunto de observações obtidas sequencialmente no tempo e tem como característica central a presença de uma dependência temporal entre elas, ou seja, são correlacionadas. Representa-se por  $\{y_t, t = 1, 2, \dots, T\}$ , onde a variável de interesse  $y$  pode ser univariada ou multivariada. Neste trabalho, consideram-se apenas séries univariadas. De uma maneira geral, o estudo de séries temporais envolve interesses como (Box et al., 2013):

- Compreender o mecanismo gerador da série - tentar obter razões sobre o comportamento da série, controlar sua trajetória, procurar periodicidades relevantes nos dados e etc;
- Predizer o comportamento futuro da série - executar planos a longo, médio ou curto prazo, exercendo uma função chave na tomada de decisões. Previsões também podem ser classificadas como *um-passo* ou *multi-passos* à frente, onde as previsões são necessárias para a próxima ou múltiplas observações futuras, respectivamente.



Formalmente, podemos definir uma série temporal  $\{y_t, t = 1, 2, \dots, T\}$  como uma realização de um processo estocástico, ou seja, uma família de variáveis aleatórias,  $\{y(t), t \in I\}$  onde em cada  $t \in I$   $y(t)$  é uma variável aleatória (v.a.) definida num espaço de probabilidades  $(\Omega, A, P)$ . O conjunto de índices temporais pode ser discreto  $I = \{1, 2, \dots, T\}$  ou contínuo  $I = \{t : t_1 < t < t_2\}$  dando origem a processos em tempo discreto ou contínuo, respectivamente.

Num processo estocástico  $\{Y_t : t = 0, \pm 1, \pm 2, \dots\}$  a função média é definida por

$$\mu_t = E(y_t) \quad (2.1)$$

a função de autocovariância (*facv*) por

$$\gamma(s, t) = cov(y_s, y_t) = E[(y_s - \mu_s)(y_t - \mu_t)] \quad (2.2)$$

e a função de autocorrelação (*fac*) representada por

$$\rho_{(s,t)} = corr(y_s, y_t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s,s)\gamma(t,t)}} \quad (2.3)$$

A estacionariedade é uma suposição importante para um processo estocástico, uma ideia básica de estacionariedade é que as leis de probabilidade que atuam no processo não mudam com o tempo, isto é, o processo mantém o equilíbrio estatístico.

Dizemos que um processo estocástico  $Y(t) = \{Y(t), t \in T\}$  é *fracamente estacionário* ou *estacionário de segunda ordem* apenas se:

- (I)  $E[Y(t)] = \mu(t) = \mu$ , constante, para todo  $t \in T$ ;
- (II)  $var[Y(t)] = E[(Y(t) - \mu)^2] = \sigma^2$ , para todo  $t \in T$ ;
- (III)  $\gamma(s, t) = Cov[Y(s), Y(t)]$  é uma função de  $|s - t|$ .

Um processo estocástico  $Y(t) = \{Y(t), t \in T\}$  diz-se Gaussiano (ou normal) se para qualquer conjunto  $\{t_1, t_2, \dots, t_n\}$  de  $T$ , as v.a.  $y(t_1), \dots, y(t_n)$  tem distribuição normal multivariada. Se um processo for Gaussiano ele será determinado pelas médias e covariâncias; em particular, se ele for *estacionário de segunda ordem*, ele será *estritamente estacionário*.

Devido a *facv*  $\gamma(s, t)$ , de uma série temporal estacionária,  $Y(t)$ , depender somente de suas diferenças  $|s - t|$ , podemos simplificar a notação, com  $|s = t + h|$ , onde  $h$  representa um *lag* (ou alterações no tempo). Logo,  $\gamma(t + h, t) = cov(y_{t+h}, y_t) = cov(y_t, y_0) = \gamma(h, 0)$ , porque a diferença entre  $t + h$  e  $t$  é a mesma que a diferença do tempo  $h$  e 0. Contudo, a *facv* de uma série temporal estacionária, pode ser reescrita para

$$\gamma(h) = cov(y_{t+h}, y_t) = E[(y_{t+h} - \mu)(y_t - \mu)] \quad (2.4)$$

e a *fac* pode ser reescrita para

$$\rho(h) = \frac{\gamma(t+h, t)}{\sqrt{\gamma(t+h, t+h)\gamma(t, t)}} = \frac{\gamma(h)}{\gamma(0)} \quad (2.5)$$

As propriedades gerais da *facv* para um processo estacionário são:

- (I)  $\gamma(0) = \text{var}(Y_t), \rho_0 = 1;$
- (II)  $\gamma(h) = \gamma(-h), \rho(h) = \rho(-h);$
- (III)  $|\gamma(h)| \leq \gamma(0), |\rho(h)| \leq 1.$

Um processo  $\{\epsilon_t : t = 0, \pm 1, \pm 2, \dots\}$  diz-se que constitui ruído branco quando é formado por uma sucessão de v.a. com a mesma distribuição, média constante  $E\{\epsilon_t\} = \mu$  (usualmente  $\mu = 0$ ), variância constante  $\text{Var}\{\epsilon_t\} = \sigma^2$  e com funções covariância e correlação iguais a:

$$\gamma_k = \begin{cases} \sigma^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \quad \rho_k = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

## Decomposição de Séries Temporais

Uma série temporal exhibe padrões ao longo do tempo nomeadamente tendência, sazonalidade e ciclo. Os métodos de decomposição identificam esses padrões individualmente e podem ser extraídos da série para uma melhor compreensão ou apenas para melhorar as previsões.

A componente tendência ( $T_t$ ) para o período  $t$  pode ser descrita quando se verifica uma alteração a longo prazo nos valores observados. O termo *trend-cycle* ( $E_t$ ) é utilizado quando a série temporal exhibe flutuações que não estão num período fixo. Componente sazonal, ( $S_t$ ) para o período  $t$  é descrita quando a série está influenciada pelos fatores sazonais, por exemplo, o trimestre do ano, o mês, ou dia da semana. A sazonalidade é sempre de um período fixo e conhecido (Hyndman and Athanasopoulos, 2014). A decomposição final poderá assumir dois modelos: aditivo e multiplicativo, representados pelas equações (2.6) e (2.7).

$$Y_t = S_t + T_t + E_t \quad (2.6)$$

$$Y_t = S_t \cdot T_t \cdot E_t \quad (2.7)$$

O modelo aditivo é mais adequado quando a magnitude das flutuações sazonais ou a variação em torno do  $E_t$  não variam com o nível da série temporal. Quando a variação no padrão  $S_t$ , ou a variação em torno do  $E_t$ , parece ser proporcional ao nível das séries temporais, um modelo multiplicativo é mais adequado. Com as séries temporais económicas, modelos multiplicativos são comuns (Hyndman and

Athanasopoulos, 2014). Note-se que é sempre possível se obter um modelo aditivo a partir de um multiplicativo pela aplicação da função logaritmo a esse modelo. Resultando o seguinte modelo na equação (2.8):

$$\log(Y_t) = \log(S_t) + \log(T_t) + \log(E_t) \quad (2.8)$$

Vários métodos são propostos na literatura para calcular as diferentes componentes da decomposição. O método de decomposição clássica teve origem na década de 1920 e é um procedimento relativamente simples, servindo de base para a maioria dos outros métodos de decomposição (Hyndman and Athanasopoulos, 2014). Outro método muito comum e robusto é o STL (do inglês, *Seasonal and Trend decomposition using Loess*) desenvolvido por Cleveland et al. (1990). Recentemente, De Livera et al. (2011) consideraram um modelo em espaço de estados que permitem modelar sazonalidades múltiplas.

## 2.2 Modelos

Os modelos de regressão são processos matemáticos que relacionam o comportamento de uma variável  $Y$  com outra  $X$ , podendo ser  $X$  um conjunto de variáveis. Quando a função  $f$  que relaciona duas variáveis é do tipo  $f(X) = a + bX$  temos o modelo de regressão simples. A variável  $X$  diz-se independente, enquanto  $Y = f(X)$  é a variável dependente. O modelo de regressão é multivariado quando envolve uma relação com mais de duas variáveis. Estes modelos são construídos com os objetivos:

- (I) **Seleção de variáveis** - Frequentemente, não se tem ideia de quais são as variáveis que afetam significativamente a variação de  $Y$ . A análise de regressão pode auxiliar no processo para seleção de variáveis eliminando aquelas cuja contribuição não seja importante;
- (II) **Estimação de parâmetros** - Dado o modelo e um conjunto de dados referente às variáveis resposta e preditoras, estimar parâmetros ou ajustar um modelo aos dados significa obter valores ou estimativas para os parâmetros, por algum processo, tendo por base o modelo e os dados observados;
- (III) **Inferência** - O ajuste do modelo de regressão em geral tem por objetivos básicos, além de estimar os parâmetros, realizar inferências sobre eles, tais como, testes de hipóteses e intervalos de confiança.

Um modelo é uma representação matemática do mundo real. Os modelos matemáticos podem ser classificados de acordo com o nível de conhecimento *a priori* utilizado na seleção de sua estrutura (Ljung, 1987):

- modelos caixa-branca (*white-box*) - tem suas estruturas completamente ajustadas a partir de informações *a priori*. A função matemática que descreve o comportamento dinâmico do sistema original é previamente conhecida;
- modelos caixa-cinza (*gray-box*) - são identificados utilizando algum conhecimento *a priori* para simplificar os algoritmos de seleção de estrutura;
- modelos caixa-preta (*black-box*) - não utiliza nenhuma informação conhecida *a priori*. A estrutura destes modelos pode ser ajustada dentro de famílias conhecidas por apresentar boa flexibilidade na modelagem de sistemas.

Os modelos de séries temporais encontradas na literatura podem ser organizados da seguinte forma: lineares e não-lineares. Os modelos lineares, designados por vezes, modelos estatísticos (De Gooijer and Hyndman, 2006) são basicamente os Autoregressivos (AR-*autoregressive*), de Médias Móveis (MA-*moving-average*), Autoregressivos de Médias Móveis (ARMA-*autoregressive moving average*) e Autoregressivos de Médias Móveis Integrados (ARIMA-*auto regressive integrated moving average*).

Há muitos modelos não lineares na literatura estatística e um modelo amplamente utilizado na modelagem de séries foi desenvolvido por Bollerslev (1986) com o objetivo de representar adequadamente a heterocedasticidade observada em séries financeiras como ilustra na Figura 2.1: o modelo GARCH.

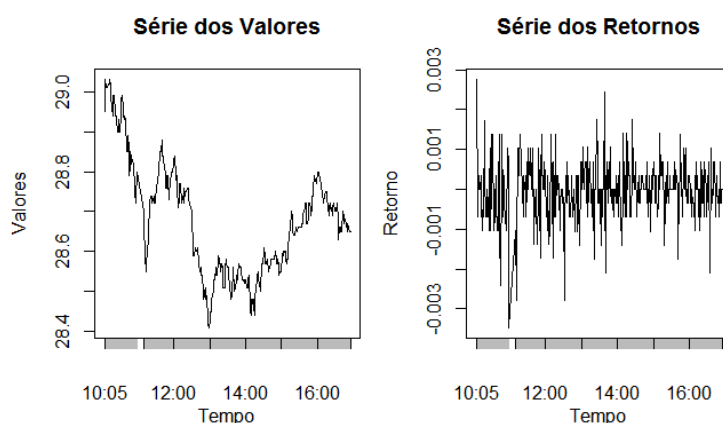


Figura 2.1: Exemplo de séries temporais (Valores e Retornos) com as cotações das ações de uma empresa em minutos.

Nas duas últimas décadas os modelos de aprendizagem automática, também conhecidos como modelos *black-box*, tem chamado a atenção de pesquisadores na comunidade de previsão. Em alguns casos de estudo, esses modelos apresentaram bons desempenho quando comparados com outros modelos gerados pela estatística

clássica. Uma vantagem a ser considerada é que modelos gerados a partir dos algoritmos de aprendizagem automática, correspondem essencialmente a modelos lineares ou não-lineares.

Werbos (1974, 1988), constatou que as redes neuronais artificiais (RNAs) obtiveram bons resultados que os métodos estatísticos clássicos, com as abordagens *Box-Jenkins*, na aplicação de um modelo para o mercado de gás natural. Num estudo similar Lapedes and Farber (1987) concluíram que as RNAs podem ser utilizadas com sucesso na modelagem e previsão de séries temporais não lineares. Mais tarde, aparecem outros modelos gerados pelos algoritmos: árvores de decisão, *support vector machine* e o k-vizinhos mais próximos (Hastie et al., 2005; Alpaydin, 2004).

### 2.2.1 Modelos Estatísticos

Dentre os modelos estacionários, os modelos AR de ordem  $p$  (AR( $p$ )) são bastantes comuns. Diz-se que o processo  $Z_t$  satisfaz um modelo AR( $p$ ) quando a solução da equação estocástica as diferenças:

$$\begin{aligned} Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - \dots - \phi_p Z_{t-p} &= \epsilon_t, \text{ ou} \\ (1 - \phi_1 \beta - \phi_2 \beta^2 - \dots - \phi_p \beta^p) Z_t &= \epsilon_t, \text{ ou} \\ \phi_p(\beta) Z_t &= \epsilon_t \end{aligned}$$

onde,  $\phi_p(\beta) = 1 - \phi_1 \beta - \phi_2 \beta^2 - \dots - \phi_p \beta^p$ , é o polinómio AR( $p$ ), associado ao processo  $Z_t$ . O processo AR( $p$ ) é sempre invertível e é estacionário quando as raízes do polinómio autoregressivo associado estão fora do círculo unitário. A *fac* deste tipo de processos apresenta um decaimento exponencial e/ou sinusoidal para zero, e a *facv* é zero a partir de  $k = p + 1$ .

Outro modelo popular é o MA( $q$ ) e juntamente com o AR( $p$ ) são úteis para definir uma nova classe de modelos, chamado os modelos combinados ARMA( $p, q$ ). O ARMA pode ser melhor alternativa para ambos modelos AR( $p$ ) e MA( $q$ ), uma vez que pode modelar uma série temporal com um modelo mais parcimonioso que induziria melhores estimativas dos parâmetros (Ming et al., 2014).

A maioria dos procedimentos de análise estatística em séries temporais supõe que estas sejam estacionárias, portanto, será necessário transformar os dados originais se estes não formam uma série estacionária. A transformação mais comum consiste em tomar diferenças sucessivas da séries original, até se obter uma série estacionária (Hyndman and Athanasopoulos, 2014). A primeira diferença  $Z_t$  é definida por:

$$\Delta Z(t) = Z(t) - Z(t - 1)$$

logo, a segunda diferença é

$$\begin{aligned}\Delta^2 Z(t) &= \Delta[\Delta Z(t)] = \Delta[Z(t) - Z(t-1)] \\ &= Z(t) - 2Z(t-1) + Z(t-2)\end{aligned}$$

De modo geral, a  $n$ -ésima diferença de  $Z_t$  é  $\Delta^n Z(t) = \Delta[\Delta^{n-1} Z(t)]$ .

O modelo  $\text{ARIMA}(p, d, q)$  é o modelo mais popular para séries temporais não estacionárias. Considerado um dos principais modelos para previsão de séries temporais, também conhecido como modelo *Box-Jenkins* (Box et al., 2013). Um modelo  $\text{ARIMA}(p, d, q)$  é dado quando o processo  $Z_t$  satisfaz a seguinte equação (2.9):

$$\phi_p(\beta)(1 - \beta)^d Z_t = \theta_q(\beta)\epsilon_t \quad (2.9)$$

onde,  $\{\epsilon_t\} \sim RB(0, \sigma^2)$ ,  $\phi_p(\beta)$  é o polinómio autoregressivo de ordem  $p$ ,  $\theta_q(\beta)$  é o polinómio de médias móveis de ordem  $q$ ,  $d$  é a ordem da diferenciação simples e tem-se que  $|\phi_i| < 1 \forall i = 1, 2, \dots, p$ ,  $|\theta_i| < 1 \forall i = 1, 2, \dots, q$  e  $d \in \mathbb{Z}^+$ .

Os modelos  $\text{ARIMA}$  exploram a autocorrelação entre os valores da série em instantes sucessivos, mas quando os dados são observados em períodos inferiores a um ano, a série também pode apresentar uma autocorrelação sazonal. Os modelos que contemplam as séries com esse comportamento são conhecidos como  $\text{SARIMA}$ . Os modelos  $\text{SARIMA}$  contêm uma parte não sazonal, com parâmetros  $(p, d, q)$  e uma sazonal com parâmetros  $(P, D, Q)_m$ , onde  $m$  representa o número de períodos sazonais. Por exemplo, um modelo  $\text{ARIMA}(1, 1, 1)(1, 1, 1)_4$  trimestral ( $m = 4$ ) pode ser escrito como:

$$(1 - \phi_1\beta)(1 - \Phi_1\beta^4)(1 - \beta)(1 - \beta^4)Z_t = (1 + \theta_1\beta)(1 + \Theta_1\beta^4)\epsilon_t$$

em que:

- $(1 - \phi_1\beta)$  representa a parte AR não sazonal de ordem 1;
- $(1 - \Phi_1\beta^4)$  representa a parte AR sazonal de ordem 1 e período sazonal 4;
- $(1 - \beta)$  é a diferenciação não sazonal de ordem 1;
- $(1 - \beta^4)$  é a diferenciação sazonal de ordem 1 e período sazonal 4;
- $(1 + \theta_1\beta)$  é a parte não sazonal de MA de ordem 1;
- $(1 + \Theta_1\beta^4)$  é a parte sazonal de MA de ordem 1 e período sazonal 4.

Para séries temporais que apresentam comportamentos heterocedásticos, Engle (1982) introduziu o modelo ARCH, do inglês, *Autoregressive Conditional Heteroscedasticity* que foi posteriormente estendido por Bollerslev (1986), dando origem

ao modelo GARCH (*Generalized ARCH*). A vantagem deste último é que pode ser utilizado para descrever a volatilidade com menos parâmetros do que um modelo ARCH e por isso é mais parcimonioso.

Um processo  $\epsilon_t$  diz-se GARCH  $\epsilon_t$  se é dado pela seguinte equação (2.10) :

$$\epsilon_t = a_t \sqrt{h_t} \quad \text{com} \quad h_t = \alpha_0 + \sum_{i=1}^r \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^s \beta_j h_{t-j} \quad (2.10)$$

em que  $\{a_t\}$  é uma sequência de variáveis aleatórias i.i.d. com média zero e variância um, em que se tem  $\alpha_0 > 0, \alpha_i \geq 0 \forall i > 0, \beta_j \geq 0 \forall j > 0, \sum_{i=1}^q (\alpha_i + \beta_i) < 1$  e  $q = \max(r, s)$ . Considerando um GARCH(1,1), temos a seguinte expressão (2.11):

$$\epsilon_t = a_t \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}} \quad (2.11)$$

com  $\alpha_0 > 0, \beta_1 \geq 0, \alpha_1 + \beta_1 < 1$ , temos  $E(\epsilon_t^2) = \alpha_0 / (1 - (\alpha_1 + \beta_1))$ .

## 2.2.2 Aprendizagem Automática

A aprendizagem automática utiliza técnicas responsáveis por descobrir padrões e relações ocultas nos dados, através do desenvolvimento de métodos computacionais que permitem aos computadores "aprender" com a experiência. O estudo de aprendizagem automática divide-se basicamente em dois grupos: supervisionada e não supervisionada. A aprendizagem supervisionada, induz conceitos a partir de exemplos que estão pré-classificados, ou seja, estão rotulados com uma classe conhecida, útil em problemas de classificação (classe possui valores discretos) e regressão (para classe de valores contínuos). Na aprendizagem não supervisionada os padrões de entrada são conhecidos e existe uma incerteza sobre a saída esperada, desta forma é necessário utilizar de métodos probabilísticos para simular uma experiência não vivida. Este trabalho, concentra-se no cenário envolvendo tarefas de aprendizagem supervisionada, com algoritmos voltados para problema de regressão. Nesta seção será considerada a principal abordagem necessária para os futuros capítulos, os algoritmos de aprendizagem automática redes neurais e o gradiente *boosting*.

### Redes Neurais

A rede do tipo *multilayer perceptron* (MLP), também conhecida por rede neuronal (RN) é um dos algoritmos mais estudados que permite modelar relações não-lineares complexas entre um conjunto de variáveis de entrada e uma variável de saída (Hyndman and Athanasopoulos, 2014). Uma rede neuronal representa uma rede de nós organizados em camadas, com: uma camada de entrada, uma ou mais camadas intermediárias (que contêm nós escondidos) e uma camada de saída, a Figura 2.2 ilustra uma típica rede MLP.

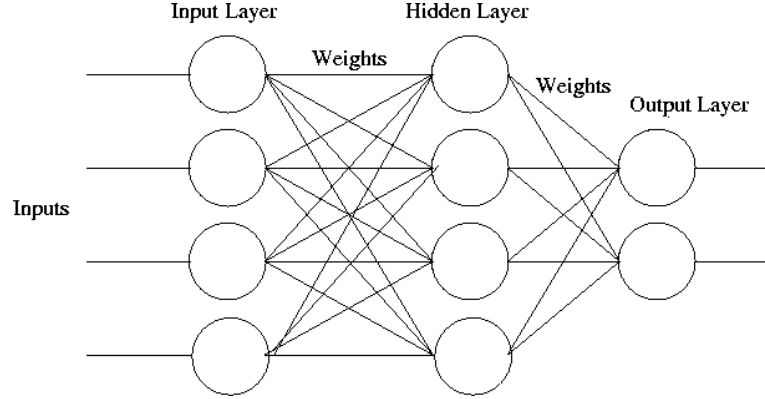


Figura 2.2: Exemplo ilustrativo de uma rede MLP

Considerando uma rede neuronal padrão com uma camada oculta dada por (Ben Taieb, 2014)

$$m(x) = \alpha_0 + \sum_{j=1}^{NH} \alpha_j g(w_j^T x') \quad (2.12)$$

onde  $x'$  representa o vector de entrada, aumentado de 1, por exemplo,  $x' = (1, x^T)^T$ ,  $w_j$  é o vector de peso para  $j$ -ésimo nó escondido,  $\alpha_0, \alpha_1, \dots, \alpha_n$  são os pesos para o nó de saída e  $NH$  é o número de nós escondidos. A função  $g$  é a saída do nó oculto e é geralmente dada pela função logística:  $g(u) = \frac{1}{1+e^{-u}}$ . O número de nós escondidos ( $NH$ ) controla a complexidade do modelo. Em particular, quando  $NH = 0$  o MLP reduz-se ao modelo linear (LIN). Para um determinado número de nós escondidos, os pesos podem ser estimados usando o algoritmo de otimização *back-propagation*. Normalmente, os pesos iniciais são escolhidos para serem de valores aleatórios próximos de zero e o algoritmo *back-propagation* atualiza os pesos de modo que os erros da previsão sejam minimizados (Ming et al., 2014).

As redes MLP constituem um modelo fortemente parametrizado e com a probabilidade de *overfitting* os dados (ver Secção (2.3)). Um método de regularização muito eficaz é reduzir a dimensão do espaço de parâmetros (dimensão da rede) ou reduzir o tamanho efetivo da dimensão de cada parâmetro (valores do decaimento de pesos) (Sarle, 1995). Finalmente, a aplicação do modelo MLP neste trabalho baseia-se no pacote *nnet* do R Core Team (2014).

### Gradiente *Boosting*

*Boosting* é uma técnica que combina a saída obtida por vários algoritmos "base" para produzir um comitê de preditores. Precisamente um algoritmo iterativo que



consiste em efetuar repetidas execuções de um algoritmo básico, modificando a distribuição de pesos no conjunto de treino e combinando os preditores obtidos, num único preditor eficiente (Gama et al., 2012). A primeira proposta foi apresentada por Schapire (1990) para resolver problemas com classificação binária. Em 1996, Freund et al. (1996), propuseram a primeira versão do algoritmo *boosting* chamado de *Adaboost* (do inglês, *Adaptive Boosting*). Esta é uma técnica muito utilizada por pesquisadores para melhorar a precisão dos algoritmos em problemas de classificação e algumas extensões do *Adaboost* incluem problemas de regressão como o *AdaBoost.R2*, introduzido por Drucker (1997), o *L2Boost*, proposto por Bühlmann and Yu (2003) e o *AdaBoost.RT* criado por Solomatine and Shrestha (2004). Uma das principais diferenças entre os algoritmos *boosting* para regressão está no cálculo da função de perda: a mais comum é a função do erro quadrático ( $L_2$ ), outra que apresenta algumas propriedades de robustez é a do erro absoluto ( $L_1$ ) (Bühlmann and van de Geer, 2011). O gradiente *boosting* foi derivado de métodos *boosting* (Freund and Schapire (1997), Friedman et al. (2000), Friedman (2001)) para estabelecer uma conexão com modelos estatísticos (Natekin and Knoll, 2013) e o objetivo desta técnica é estimar a função de regressão ótima para  $f(*)$ , definida por:

$$\hat{f}(*) = \underbrace{\arg \min}_{f(*)} \{E_{Y,X}[\rho(Y, f(X))]\} \quad (2.13)$$

onde  $\rho$  representa a função de perda. Por exemplo, esta função de perda poderia ser definida pela  $L_2$ ,  $\rho(y, f(*)) = (y - f)^2$  o que conduz à regressão clássica dos mínimos quadrados com  $f(x) = E(Y|X = x)$ . Em alternativa pode considerar-se para função perda a  $L_1$ ,  $\rho(y, f(*)) = |y - f|$  o que conduz à regressão na mediana. O risco *empírico* é minimizado por,

$$\hat{f}(*) = \underbrace{\arg \min}_{f(*)} \left\{ \frac{1}{n} \sum_{i=1}^n \rho(y_i, f(x_i)) \right\} \quad (2.14)$$

Em cada iteração *boosting*  $j$ , o *base-learner* ajusta portanto, os erros cometidos na iteração anterior  $y - f(*)^{[j-1]}$ .

A ideia principal do gradiente *boosting* é ajustar o *base-learner*, não para reajuste dos pesos observados, mas para o vector gradiente negativo  $u^{[j]}$  da função de perda  $\rho(y, \hat{f}(x))$  avaliado na iteração anterior  $j - 1$ :

$$\mathbf{u}^{[j]} = (u_i^{[j]})_{i=1, \dots, n} = \left( - \frac{\partial}{\partial f} \rho(y_i, f) \bigg|_{f=\hat{f}^{[j-1]}} \right)_{i=1, \dots, n} \quad (2.15)$$

O principal objetivo do algoritmo é minimizar (2.14) para a função (2.13) e para tal, pode ser utilizado no algoritmo *component-wise gradient boosting* (Hofner et al., 2014):

1. Inicializar a função  $\hat{f}^{[0]}$ . De notar que  $\hat{f}^{[0]}$  é um vetor de comprimento  $N$ . Nos parágrafos seguintes  $\hat{f}^{[j]}$  refere-se ao vetor estimado na iteração  $j$ .
2. Especificar um conjunto de *base-learners*. *Base-learners* são estimadores simples de regressão com um conjunto fixo de variáveis de entrada e uma resposta univariada. O conjunto de variáveis de entrada pode variar nos diferentes *base-learners* e normalmente são subconjuntos de pequena dimensão das variáveis explanatórias. Seja  $P$  o número de *base-learners* e defina-se  $j = 0$ .
3. Incrementar  $j$  por 1, sendo  $j$  o número de iterações.
4. (a) Calcular o gradiente negativo  $-\frac{\partial}{\partial \mathbf{f}}$  da função de perda e avaliá-lo em  $\hat{f}^{[j-1]}(x_i) i = 1, \dots, n$ , o que produz o vetor gradiente negativo de (2.15);  
 (b) Estimar os coeficientes dos  $P$  *base-learners*: isto é, usar cada um dos estimadores de regressão definidos no passo 2 separadamente para obter  $P$  vetores, onde cada um é um estimador do vetor gradiente nulo  $\mathbf{u}^{[j]}$ .  
 (c) Selecionar o *base-learner* que melhor se ajusta a  $\mathbf{u}^{[j]}$  segundo o critério da soma dos quadrados dos resíduos e definir  $\hat{\mathbf{u}}^{[j]}$  com os valores desse *base-learner*.  
 (d) Atualizar a estimativa atual definindo  $\hat{f}^{[j]} = \hat{f}^{[j-1]} + v\hat{\mathbf{u}}^{[j]}$ , onde  $0 < v \leq 1$  é o parâmetro que controla o passo do gradiente.
5. Repetir passos 3 e 4 até que o critério de paragem seja atingido.

O procedimento *boosting* depende de dois hiperparâmetros importantes: o factor *shrinkage* ( $v$ ) e o número de componentes (ou número de iterações *boosting* ( $J$ )). O valor de  $v$  afeta o melhor valor para  $J$ , por exemplo, diminuindo o valor de  $v$  requer aumentar o valor de  $J$ . Uma vez que eles controlam o grau de ajuste, devemos encontrar o melhor valor para ambos minimizando algum critério de seleção do modelo. Friedman (2001), mostra que pequenos valores de  $v$  são melhores em termos de menos *overfitting* do procedimento *boosting*.

Vários *base-learners* têm sido utilizados na literatura de *boosting*. Neste trabalho, será utilizado a aplicação do *base-learner* com *smooth effect* (por exemplo, funções da covariável que são necessárias para ser suficientemente suave, dois exemplos de estimativas das funções *smooth* são dadas na Figura 2.3) baseado em *P-splines* (do inglês, *Penalised regression splines*) univariado e bivariado, como utilizado em Schmid and Hothorn (2008) e Ben Taieb (2014). Será denotado como BST-L2 (com a função de perda  $L_2$ ) e BST-L1 (com a função de perda  $L_1$ ). Para *P-splines* univariado o *base-learner* depende apenas de uma variável de entrada, já no bivariado cada *base-learner* dependerá de um par de variáveis mais outras classes *smooth effect*: o número de *knots*, o parâmetro *smoothing* e os graus de liberdade ( $gl$ ).

Finalmente, para a implementação do algoritmo gradiente *boosting* será utilizado o pacote *mboost* (Hofner et al., 2014) do R Core Team (2014).

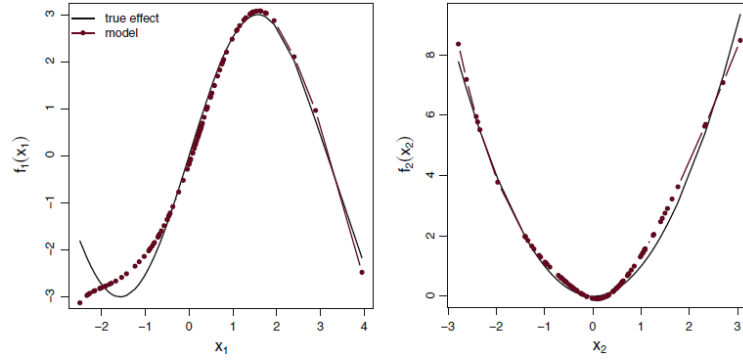


Figura 2.3: Exemplo para estimação *smooth effect* usando apenas a variável  $x_1$  (esquerda) e a variável  $x_2$  com classes *smooth effect* ( $knots = \text{quantile}(x_2, c(0.25, 0.5, 0.75))$ ,  $df=5$ ) (direita). Assim, o *P-splines* da esquerda usa por defeito o valor de 20 *knots* e 4 graus de liberdade, enquanto o *P-splines* estimado da direita é derivado de 3 *knots* colocados nos quartis e 5 graus de liberdade. Fonte Hofner et al. (2014).

## 2.3 Seleção de Modelos

Quando um modelo perde a capacidade de generalizar ocorre um fenômeno denominado de sobre ajuste (*overfitting*), onde a complexidade da função  $f$  obtida é superior à necessidade do problema. Além deste existe outro problema que ocorre de forma contrária, chamado de sub-ajuste (*underfitting*), onde a complexidade da função  $f$  obtida é inferior a necessidade do problema (Gama et al., 2012). A Figura 2.4 ilustra a ocorrência desses fenômenos.

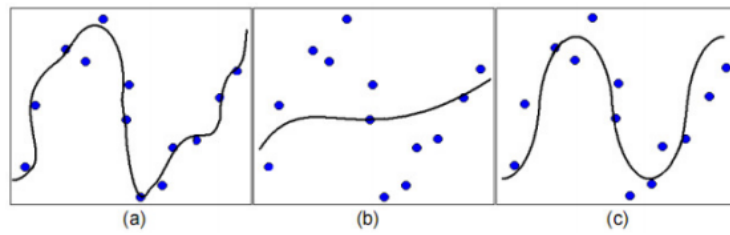


Figura 2.4: Ilustração da ocorrência de: (a) sobre-ajuste, (b) sub-ajuste, (c) função de aproximação mais adequada.

Para que a generalização do modelo não sofra de sobre-ajuste ou sub-ajuste, é necessário haver um equilíbrio entre o viés e variância, obtido através de seu dimensionamento. Existem dois métodos que abordam *overfitting*: regularização e validação (Abu-Mostafa et al., 2012). A regularização consiste em penalizar o erro dentro da

amostra por um termo *overfit penalty* que representa uma estimativa do optimismo ou a polarização do erro dentro da amostra em relação ao erro fora da amostra. A validação, por outro lado, calcula diretamente os erros fora da amostra, utilizando métodos de reamostragem; divide o conjunto de dados ( $D$ ) em treinamento ( $D_{train}$ ) e validação ( $D_{val}$ ) e dessa forma, é possível utilizar o conjunto de treinamento para atualização dos parâmetros e selecionar variáveis, e o de validação para estimar a capacidade de generalização do modelo. Existem outras abordagens propostas na literatura para calcular termo *overfit penalty*, como *Akaike Information Criterion* (AIC) (Akaike, 1969) e *Bayesian Information Criterion* (BIC) (Schwarz, 1978).

Em vez de considerar uma única divisão para  $D$ , também é possível obter múltiplas divisões e combinar seus resultados através do método de validação cruzada (do inglês, *cross-validation*). Uma variante muito divulgada na literatura é o *k-fold cross-validation* que divide o  $D$  em  $k$  partes  $D_1, \dots, D_k$ , onde  $D_{val}$  tem um tamanho de  $N/k$ . Para todos esses conjuntos são treinadas a hipótese sobre  $D/D_{val}$  e calculado o erro da validação em  $D_{val}$ . Um *trade-off* entre viés e variância depende do valor de  $k$ . Na prática, foi mostrado empiricamente que estes valores  $k = 5$  ou  $k = 10$  proporcionam uma bom *trade-off* entre viés-variância (Kohavi et al., 1995).

Hyndman (2014), recomenda o uso de validação cruzada em séries temporais, sempre que possível. Nesta, abordagem o autor propõe usar vários conjuntos de treinamento diferentes, cada um contendo uma observação a mais que a anterior. A Figura 2.5, mostra uma série com conjunto de treino (em azul) e outro de teste (em vermelho) e as medidas de precisão da previsão (ver Secção 2.4) são calculadas em cada conjunto de teste.

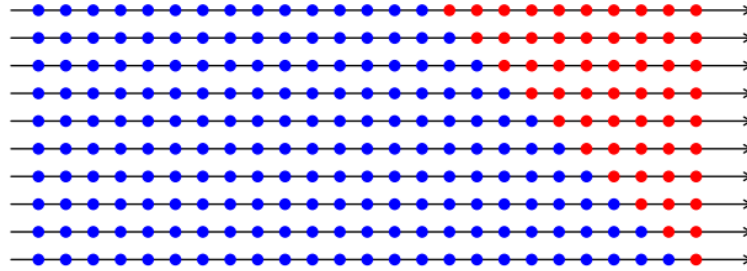


Figura 2.5: Validação cruzada de séries temporais, onde cada conjunto de treino (azul) contém uma observação a mais que o conjunto anterior, e consequentemente cada conjunto de teste (vermelho) tem menos observações que o anterior. Fonte (Hyndman, 2014).

Uma variação dessa abordagem se concentra em um único horizonte de previsão para cada conjunto de teste. A Figura 2.6, ilustra uma série com conjuntos de teste que contêm apenas uma observação em cada caso.

Em qualquer uma dessas abordagens de validação cruzada, precisamos de um tamanho mínimo para o conjunto de treinamento porque muitas vezes não é possível

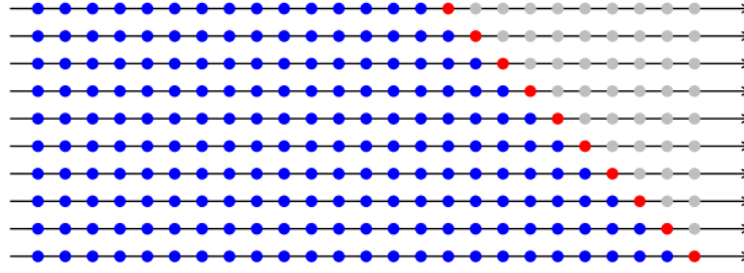


Figura 2.6: Validação cruzada de séries temporais com base em previsões de um-passo. Os pontos azuis representam os conjuntos de treino, os vermelhos de teste e os pontos cinzentos são ignorados. Fonte (Hyndman, 2014).

fazer qualquer previsão significativa se não houver dados suficientes para estimar o nosso modelo escolhido. O tamanho mínimo do conjunto de treinamento depende da complexidade do modelo que deseja usar (Hyndman, 2014).

Suponha que  $k$  observações são necessárias para produzir uma previsão confiável  $h$ -passos à frente. O procedimento é o seguinte:

1. Para  $i = 1, 2, \dots, T - k - h + 1$ , onde  $T$  é o número total de observações, selecionar a observação no tempo  $k + h + i - 1$  para o conjunto de teste, e usar as observações do tempo  $1, 2, \dots, k + i - 1$  para estimar o modelo de previsão. Calcular o erro  $h$ -passos da previsão para o tempo  $k + h + i - 1$ ;
2. Calcular as medidas de precisão das previsões com base nos erros obtidos.

Quando  $h = 1$ , isto dá o mesmo procedimento tal como descrito acima.

## 2.4 Métricas de Avaliação

Para avaliação dos diferentes métodos de previsão é necessário comparar o valor  $y_i$  conhecido e o valor predito pelo modelo, ou seja,  $\hat{f}(x_i)$ . Muitas medidas de erro de precisão  $e_i = y_i - \hat{f}(x_i)$  estão propostas na literatura para avaliar o desempenho dos métodos de previsão (Hyndman and Athanasopoulos, 2014).

- **Erro Dependente da Escala** - As medidas de precisão que são baseadas em  $e_i$ , são, dependentes de escala e não podem ser utilizadas para fazer comparações entre as séries que estão em diferentes escalas. As duas medidas utilizados são o MAE (*mean absolute error*) e RMSE (*root mean squared error*):

$$MAE(\hat{f}) = \text{mean}(|e_i|) \quad (2.16)$$

$$RMSE(\hat{f}) = \sqrt{\text{mean}(e_i^2)} \quad (2.17)$$

- **Erro Percentual** - é dado por  $p_i = 100 \frac{e_i}{y_i}$ . Possui a vantagem de ser independente da escala e, assim é frequentemente usado para comparar o desempenho dos métodos de previsão entre diferentes conjuntos de dados. A medida utilizada é o MAPE (*mean absolute percentage error*):

$$MAPE(\hat{f}) = \text{mean}(|p_i|) \quad (2.18)$$

- **Erro Escalado** - proposto por [Hyndman and Koehler \(2006\)](#), como uma alternativa ao uso dos erros percentuais quando comparado a precisão das previsões em toda série em diferentes escalas. A medida MASE (*mean absolute scaled error*) é dada por:

$$MASE(\hat{f}) = \text{mean}(|q_j|) \quad (2.19)$$

onde

$$q_j = \frac{e_j}{\frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}|}$$

## 2.5 Previsão Multi-Passos

Perante um problema de previsão multi-passos, temos que escolher a melhor estratégia de previsão a utilizar. A literatura distingue duas estratégias de previsões multi-passos: recursiva e direta. Grande parte das pesquisas tem-se centrado numa comparação competitiva destas duas estratégias. No entanto [Ben Taieb \(2014\)](#), propõe a estratégia de retificação que combina as melhores propriedades de ambas estratégias direta e recursiva. Essas técnicas estão descritas brevemente nas subseções a seguir.

A principal tarefa da previsão multi-passos em séries temporais univariadas com  $T$  observações, consiste em estimar as observações futuras de  $y_1, \dots, y_{T+H}$ , em que  $h = 1, \dots, H$  representa o horizonte da previsão, podendo ser tipicamente classificados em curto, médio ou longo prazo.

Assumindo que a série temporal  $y_1, \dots, y_T$  é a realização de um processo autoregressivo na forma de:

$$y_t = f(\mathbf{x}_{t-1}) + \epsilon_t \quad \text{com} \quad \mathbf{x}_{t-1} = [y_{t-1}, \dots, y_{t-d}]' \quad (2.20)$$

que é especificada por uma função  $f$ ,  $d$  é a ordem do modelo e um termo de ruído  $\{\epsilon_t\}$  que é um processo estocástico de ruído branco com  $\mathbb{E}[\epsilon_t] = 0$  e  $\mathbb{E}[\epsilon_t^2] = \sigma_\epsilon^2$ . Considerando o MSE a medida de erro a ser minimizada, então, a previsão ótima

no horizonte  $h = 1, \dots, H$  é a média condicional  $\mu_{t+h|t} = \mathbb{E}[y_{t+h}|x_t]$  e que o objetivo da previsão é estimá-la.

### 2.5.1 Estratégia Recursiva

A estratégia recursiva concentra-se na construção de um único modelo ( $f$ ), com o objetivo de minimizar a variância do erro de previsão um-passo à frente. Em palavras outras, isso implica um modelo na forma de (Ben Taieb, 2014)

$$y_t = m(\mathbf{z}_t; \phi) + e_t \quad (2.21)$$

onde  $\mathbf{z}_{t-1} = [y_{t-1}, \dots, y_{t-p}]'$ ,  $p$  é uma estimativa verdadeira da ordem do modelo  $d$  (por exemplo,  $p = \hat{d}$ ),  $\phi = [\Psi, \beta]$  são os parâmetros do modelo onde  $\Psi$  é um conjunto de hiperparâmetros e  $\beta$  é um conjunto de parâmetros, e  $e_t = f(\mathbf{x}_{t-1}) - m(\mathbf{z}_{t-1}; \phi) + \epsilon_t$  é o erro da previsão do modelo  $m$  com  $\mathbb{E}[e_t] = 0$ .

A ordem do modelo e os hiperparâmetros são estimados pela minimização dos erros um-passo à frente, como segue

$$(p, \hat{\Psi}) = \underbrace{\operatorname{argmin}}_{p, \Psi} \sum_{(\mathbf{z}_{t-1}, y_t) \in D_{val}} [y_t - m(\mathbf{z}_{t-1}; \Psi, \hat{\beta})]^2 \quad (2.22)$$

onde  $D_{val}$  é o conjunto de validação e  $\hat{\beta}$  é o parâmetro estimado usando o conjunto de treino  $D_{train}$ , como descrito na Secção (2.3). Mais precisamente, as previsões  $\hat{\mu}_{T+h|T}$ , são calculadas da seguinte forma

$$m^{(h)}(\mathbf{z}_T, \hat{\phi}) = \begin{cases} m(\mathbf{z}_T, \hat{\phi}), & \text{se } h = 1, \\ m(m^{(h-1)}(\mathbf{z}_T, \hat{\phi}), \dots, m^{(1)}(\mathbf{z}_T, \hat{\phi}), y_T, \dots, y_{T-p+h}; \hat{\phi}), & \text{se } 1 < h \leq p, \\ m(m^{(h-1)}(\mathbf{z}_T, \hat{\phi}), \dots, m^{(h-p)}(\mathbf{z}_T, \hat{\phi}); \hat{\phi}), & \text{se } h > p, \end{cases} \quad (2.23)$$

Uma variação na estratégia recursiva é a utilização de um conjunto diferente de parâmetros para cada previsão horizonte:

$$(p_h, \hat{\Psi}_h) = \underbrace{\operatorname{argmin}}_{p, \Psi} \sum_{(\mathbf{z}_{t-h}, y_t) \in D_{val,h}} [y_t - m^{(h)}(\mathbf{r}_{t-h}; \Psi, \hat{\beta}_h)]^2 \quad (2.24)$$

onde  $D_{val,h}$  é o conjunto de validação para o horizonte  $h$ ,  $\hat{\beta}_h$  é o parâmetro estimado usando o conjunto de treino  $D_{train,h}$  e  $m^{(h)}$  é o modelo definido em (2.23). Uma vantagem da utilização desta estratégia é que apenas um modelo é necessário, poupando um significativo tempo computacional. Por outro lado, a grande desvantagem é que esta estratégia é sensível à acumulação de erros com o horizonte da previsão (Ben Taieb and Hyndman, 2012). Um exemplo desta estratégia está presente no Anexo (A.1).

### 2.5.2 Estratégia Direta

A estratégia direta difere da recursiva pois não utiliza a recursão no cálculo das previsões, impedindo assim a acumulação de erros introduzido na primeira. A estratégia direta adapta diferentes modelos de previsão diretamente para o horizonte em causa, mas sempre tendo como entrada os dados reais (Sorjamaa and Lendasse, 2006). Em outras palavras, um modelo de previsão diferente é utilizados para cada horizonte (Ben Taieb, 2014):

$$y_t = m_h(\mathbf{r}_{t-h}; \gamma_h) + e_{t,h} \quad (2.25)$$

onde, em contraste com a estratégia anterior,  $m_h$  é um modelo direto,  $\mathbf{r}_{t-h} = [y_{t-h}, \dots, y_{t-h-p_h}]'$ ,  $p_h$  é a ordem do modelo para o horizonte  $h$ ,  $\gamma_h = [\Psi_h, \beta_h]$  são os parâmetros do modelo do horizonte  $h$  onde  $\Psi_h$  representa um conjunto de hiperparâmetros e  $\beta_h$  um conjunto de parâmetros,  $e_{t,h}$  é o erro da previsão do modelo  $m_h$  no horizonte  $h$  e  $h = 1, \dots, H$ .

Para cada modelo  $m_h$ , a ordem do modelo e os hiperparâmetros são estimados através da minimização direta  $h$ -passos à frente, como segue:

$$(p_h, \hat{\Psi}_h) = \underbrace{\underset{p, \Psi}{\operatorname{argmin}}} \sum_{(\mathbf{r}_{t-h}, y_t) \in D_{val,h}} [y_t - m_h(\mathbf{r}_{t-h}; \Psi, \hat{\beta}_h)]^2 \quad (2.26)$$

onde  $D_{val,h}$  é o conjunto de validações para o horizonte  $h$  e  $\hat{\beta}_h$  é o parâmetro estimado usando o conjunto de treino  $D_{train}$  para o horizonte  $h$ .

A previsão multi-passos é obtida para cada horizonte do modelo correspondente,  $\mu_{T+h|T} = m_h(\mathbf{r}_T; \hat{\gamma}_h)$ . Nota-se que a cada passo de tempo um modelo é incorporado com entradas e estruturas próprias. Em princípio acredita-se no aumento de complexidade, devido ao número de modelos a serem identificados que depende do número de passos desejados para previsão, o que pode ser compensado pela ausência de erro acumulativo (Sorjamaa and Lendasse, 2006). Um exemplo desta estratégia está presente no Anexo (A.2).

### 2.5.3 Estratégia de Retificação

A ideia principal desta estratégia é produzir previsões recursivas (Secção 2.5.1) a partir de um modelo autoregressivo linear e em seguida ajustar estas previsões utilizando um modelo não linear direto (Secção 2.5.2). Taieb and Hyndman (2014) consideraram o modelo não linear gerado pelo algoritmo k-vizinhos mais próximos, mas explicam que qualquer outro modelo de aprendizagem não-linear pode ser utilizado para estimar os modelos da estratégia de retificação.

Primeiramente modela-se a série temporal usando um modelo linear autoregressivo (qualquer outro modelo pode ser utilizado para gerar o modelo linear, esta etapa é conhecida como *modelo base*), dado por



$$Y_t = \underbrace{\Phi_0 + \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p}}_{m(z_{t-1}; \phi)} + e_t \quad (2.27)$$

onde  $\Phi = [\Phi_0, \dots, \Phi_p]$  representam os parâmetros do modelo linear. Tanto esses parâmetros quanto a ordem do modelo  $p$  podem ser estimados utilizando métodos descritos na Secção 2.3. Em seguida, são produzidas previsões recursivas a partir do modelo estimado:  $m^{(h)}(z_t; \hat{\phi})$ . Esta fase pode ser considerada como uma primeira aproximação da expectativa condicional  $\mu_t + h|t$ . Uma etapa adicional é feita para ajustar as características dos dados que não foram bem representadas pelo ajuste linear autoregressivo, ou seja, o modelo  $m^{(h)}$  é ajustado com

$$\underbrace{y_t - m^{(h)}(\mathbf{z}_{t-h}; \hat{\phi})}_{\bar{y}_t} = m_h(\mathbf{r}_{t-h}; \gamma_h) + e_{t,h} \quad (2.28)$$

onde  $\bar{y}_t$  é a nova variável resposta na regressão,  $\mathbf{z}_{t-h} = [y_{t-h}, \dots, y_{t-h-p}]'$ ,  $\mathbf{r}_{t-h} = [y_{t-h}, \dots, y_{t-h-p_h}]'$ ,  $\gamma_h$  são os parâmetros do modelo (direto) retificação e  $h = 1, \dots, H$ . Em contraste com os parâmetros do modelo direto em (2.25), os parâmetros do modelo de retificação em (2.28) dependem dos valores do parâmetro  $\hat{\phi}$ . Eles podem ser estimados como em (2.26) mas com a nova variável de resposta  $\bar{y}_t$ . Em seguida, as previsões finais são obtidas para cada horizonte adicionando as retificações às previsões do modo de base:  $\hat{\mu}_T + h|T = m^{(h)}(\mathbf{z}_T; \hat{\phi}) + m_h(\mathbf{r}_T; \hat{\gamma}_h)$ . Resumindo a estratégia de retificação apresenta as seguintes fases:

- $y_1, \dots, y_T$  : série temporal com  $T$  observações;
  - $H$  : horizonte da previsão.
1. Ajustar o modelo AR( $p$ ) em (2.27) para obter  $m(\cdot; \hat{\phi})$ ;
  2. Para cada  $h \leftarrow 1, \dots, H$ , fazer:
    - (a) Calcular a previsão recursiva  $h$ -passo à frente do modelo AR( $p$ )  $m^{(h)}(z_{t-h}; \hat{\phi})$ , para obter uma primeira estimativa de  $\mu_t + h|t$ ;
    - (b) Calcular a nova variável  $\bar{y}_t$  como definido em (2.28);
    - (c) Treinar o modelo de retificação (direto)  $m_h(\mathbf{r}_{t-h}; \hat{\gamma}_h)$  como definido em (2.28) para o conjunto  $D = \{(\mathbf{r}_{t-h}, \bar{y}_t)\}_{t=1}^T$ .
  3. A previsão final é dada por  $\hat{\mu}_T + h|T = m^{(h)}(\mathbf{z}_T; \hat{\phi}) + m_h(\mathbf{r}_T; \hat{\gamma}_h)$  para  $h = 1, \dots, H$ .

De notar que a estratégia de retificação é diferente de uma combinação de previsão equi-pesadas das estratégias recursivas e direta, onde as previsões são geradas como se segue

$$\hat{\mu}_T + h|T = \frac{m^{(h)}(\mathbf{z}_T; \hat{\phi})}{2} + \frac{m_h(\mathbf{r}_T; \hat{\gamma}_h)}{2} \quad (2.29)$$

onde  $\hat{\phi}$  e  $\hat{\gamma}$  são os parâmetros presentes nos modelos (2.21) e (2.25), respectivamente. De facto, com a estratégia de retificação, os valores dos parâmetros do modelo (direto) de retificação  $\hat{\gamma}_h$  depende dos parâmetros do *modelo de base* recursivo  $\hat{\phi}$  e, portanto, os modelos de retificação (direto) são complementares ao modelo base (recursivo). No entanto, com a combinação igualmente ponderada na equação (2.29) as previsões são obtidas através de uma média simples das previsões recursivas e diretas, gerada independentemente (Ben Taieb, 2014).

Ao considerar modelos de aprendizagem automática não-lineares para a retificação, existe uma chance de *overfitting*, especialmente com a série temporais curtas em horizontes longos. Portanto, uma segunda abordagem foi considerada em Ben Taieb and Hyndman (2014), chamada estratégia de *boost* que estima os modelos de retificação utilizando o algoritmo de gradiente *boosting* (ver a Secção 2.2.2). Mais precisamente, o procedimento *boosting* é utilizado para modelar um sinal potencial deixado nos resíduos das previsões (2.24), em cada horizonte  $h$  com a estratégia direta. Para assegurar que os ajustes são *pequenos*, cada procedimento *boosting* obriga à escolha de um *base-learner*. Logo, o modelo de retificação  $m_h$  em (2.24) poderá ser escrito como uma soma de *pequenos* ajustes ou componentes *boosting* como segue:

$$\underbrace{y_t - m^{(h)}(\mathbf{z}_{t-h}; \hat{\phi})}_{\hat{y}_t} = \sum_{j=1}^{J_h} v * l^{[j]}(x_{t-h}; \Psi) + e_{t,h} \quad (2.30)$$

onde  $J_h$  é número de iterações *boosting* para o horizonte  $h$ ,  $v * l^{[j]}$  é o *base-learner* ajustado na iteração  $j$  com  $0 < v \leq 1$  sendo o factor *shrinkage*.

A estratégia de retificação *boost* depende de diferentes parâmetros: (I) o número de iterações *boosting*; (II) o *base-learner*; (III) o fator *shrinkage*. O número de iterações *boosting*  $J$  é o principal hiperparâmetro a ser selecionado, pois controla a estimativa do *trade-off* entre viés e variância, logo o método de *cross-validation* pode ser utilizado para obter a melhor iteração (Ben Taieb, 2014). Com relação aos parâmetros *knots* e *smoothing* requeridos por *P-splines* (ver na Secção 2.2.2), Ben Taieb (2014), apresenta que o número de *knots* não causam tanto efeito sobre a estimativa produzida e a debilidade do *P-splines* está medida por *gl* (graus de liberdade). Schmid and Hothorn (2008), propõe que o parâmetro *smoothing* deve ser ajustado para dar um valor pequeno para o grau de liberdade (por exemplo,  $gl \in [3, 4]$ ), e que este número deve se manter fixo em cada  $J$  iterações *boosting*.

## Capítulo 3

### Casos de Estudo

Este capítulo apresenta a aplicação dos métodos descritos anteriormente a casos de estudo constituídos por séries temporais do mundo real. Descreve todas as etapas de identificação, seleção e estimação dos modelos para gerar previsões, juntamente com os resultados obtidos. A série temporal do primeiro caso (na Secção 3.1) refere-se ao consumo horário de energia elétrica na cidade de Cabinda em Angola e o objetivo é prever o consumo de um dia, ou seja, 24h. O segundo caso (na Secção 3.2), apresenta uma série de retornos diários do *New York Stock Exchange* (NYSE) e o propósito é a previsão de uma semana, isto significa 5 dias em séries financeiras. Por fim, o terceiro e último caso está na Secção 3.3 e trata-se de uma série temporal com dados macroeconômicos referentes ao índice de produção mensal do *Federal Reserve Board*, onde o objetivo é prever os meses referentes a um ano.

Neste estudo, consideraram-se modelos estatísticos, ARIMA e GARCH e modelos gerados por algoritmos de aprendizagem automática, nomeadamente por redes neurais (MLP e LIN) e gradiente *boosting* (BST). Mais detalhe sobre esses modelos são apresentados nas Secções 2.2.1 e 2.2.2.

As previsões são geradas pelas estratégias: recursivas (REC) como definido na Secção 2.5.1 com modelos LIN, MLP, ARIMA e GARCH, diretas (DIR) apenas com LIN e MLP (ver na Secção 2.5.2) e de retificação (RTY) (ver Secção 2.5.3), onde o modelo linear recursivo será REC-LIN e o não linear direto o modelo gradiente *boosting* (BST), com funções de perda  $L_1$  e  $L_2$ , ambas produzidas com *P-splines* univariado (UN) e bivariado (BI). Esta estratégia será representada por RTY-BST- $L_1$  e RTY-BST- $L_2$  [UN, BI].

Algumas pesquisas foram executadas no âmbito de reutilizar pacotes já existentes, diminuindo o tempo de desenvolvimento. Embora o pacote *forecast* (Hyndman and Khandakar, 2007) no R Core Team (2014), permita aplicar diversos métodos de previsão não parece estar ainda disponível um pacote para aplicação de diferentes estratégias de previsão multi-passo à frente, especialmente para os algoritmos de aprendizagem automática. Nos métodos estatísticos apenas a estratégia recursiva foi avaliada, uma vez que não há nos pacotes usuais a estratégia direta associada.

Contudo, procedeu-se à implementação de funções para obter previsões recursivas e diretas geradas por algoritmos lineares e não lineares de aprendizagem automática. Para a estratégia RTY-BST desenvolveu-se um código para gerar previsões com retificação utilizando o algoritmo gradiente *boosting*, baseado no artigo de [Taieb and Hyndman \(2014\)](#) e reescrito na Seção 2.5.3 deste trabalho. Os parágrafos seguintes relatam a experiência, os métodos e pacotes aplicados para execução do capítulo.

Inicializou-se com uma análise exploratória de dados nas séries que permite compreender informações valiosas sobre o processo que gera os dados e como estão caracterizados. É importante salientar que em todos os casos as séries não apresentaram nenhuma inconsistência (mal preenchidos ou em branco).

Em seguida, foi realizada a seleção e estimação dos modelos, onde primeiramente o conjunto de dados  $D$  foi dividido em dois, um  $D_{treino}$  utilizado para construir o modelo e outro de  $D_{teste}$  para determinar o desempenho do modelo com os dados que não foram previamente utilizados. A seleção da ordem dos modelos fundamentou-se através da análise função de autocorrelação da variável objetivo no tempo e foram treinados com diferentes parâmetros de entrada.

Para o conjunto  $D_{treino}$  foi utilizado o pacote *caret* ([Kuhn, 2008](#)) do [R Core Team \(2014\)](#). As diferentes ordens do modelo e os hiperparâmetros foram estimados e avaliados pelo método de amostragem *k-fold cross validation* (mais detalhes na Seção 2.3), onde 75% do  $D_{treino}$  é utilizado para treinar e a parte restante para validação. A medida de precisão das previsões para validação é o RMSE (ver Seção 2.4). Atualmente, o pacote *caret* não suporta os modelos de séries temporais univariadas da estatística clássica, como ARIMA e GARCH.

O modelo linear de aprendizagem automática (LIN) que será utilizado é uma rede neuronal contendo o número de camadas ocultas e o peso iguais a 0 (zero), com *skip* camadas (ver Seção 2.2.2). Nos modelos BST-L1 e BST-L2 (ver Seção 2.2.2), o número do limite máximo de iterações é 500 e o fator *shrinkage* é fixo e igual a 0.3. Para *P-splines* bivariado o parâmetro *gl* variou entre [3,4] e o número de *knots* será fixo em 20, além disso foi adicionada outra classe *smoth effect* para decomposição *P-spline* (*center = TRUE*), uma solução apresentada por [Fenske et al. \(2011\)](#), uma vez que *P-spline* não pode fazer os graus de liberdade arbitrariamente pequenos.

Para finalizar uma análise dos resultados será apresentada e discutida de forma a se definir qual modelo aliado a melhor estratégia obteve o melhor desempenho.

### 3.1 Caso 1 - Consumo de Energia Elétrica

O consumo de eletricidade está influenciado por um conjunto de fatores, incluindo as condições climáticas (temperatura, humidade, velocidade do vento e precipitações), efeitos de calendário, atividades económicas, o crescimento da população e o preço da eletricidade. Normalmente esses fatores são utilizados como variáveis explicativas para gerar modelos de previsão de consumo elétrico. No entanto, para este estudo foi utilizado apenas o consumo.

Os dados referentes ao consumo de energia pertencem à Estação Elétrica de Malembo em Cabinda e foram fornecidos pela ENE (Empresa Nacional de Eletricidade em Angola). A série temporal possui valores observados referentes ao consumo de energia em *mega-watt*/horas (MWh) entre as 00:00AM do dia 01 de janeiro 2011 e as 23:00PM do dia 31 de dezembro de 2014 (3 anos de dados de hora em hora, totalizando 35064 observações).

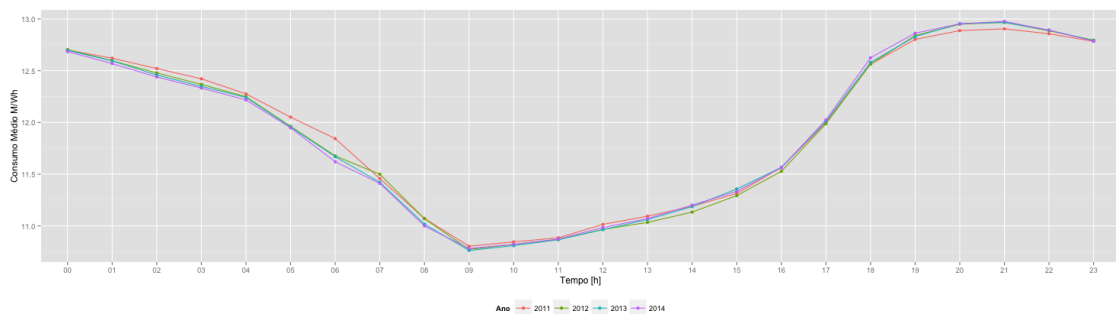


Figura 3.1: Consumo médio (MWh) horário de eletricidade no período que se estende de janeiro de 2011 à dezembro de 2014.

A Figura 3.1, apresenta o consumo médio (MWh) horário de energia elétrica, onde é possível observar que o consumo permaneceu estável entre os anos em estudo, com exceção de 2011 que apresenta um consumo maior relativamente aos anos seguintes. O padrão de consumo noturno é mais elevado que o diurno e tem a ponta da noite (17:00h) mais prolongada. Entre as 22h e 09h o consumo é decrescente e há poucas variações entre as 09h e 14h, quando volta a aumentar.

Ao observar o padrão de consumo diário na Figura 3.2 é notória uma redução do consumo sempre que ocorre uma transição entre as estações quentes para as mais frias. O inverno de 2011 apresenta um consumo mais elevado relativamente ao inverno dos anos seguintes. Ainda é possível concluir que a série não apresenta tendência aparente e tem uma sazonalidade bastante marcada, características comuns nestes tipos de séries que sofrem alterações significativas consoantes à estação do ano, horários do dia, etc., consequência natural dos hábitos e rotina diária das pessoas.

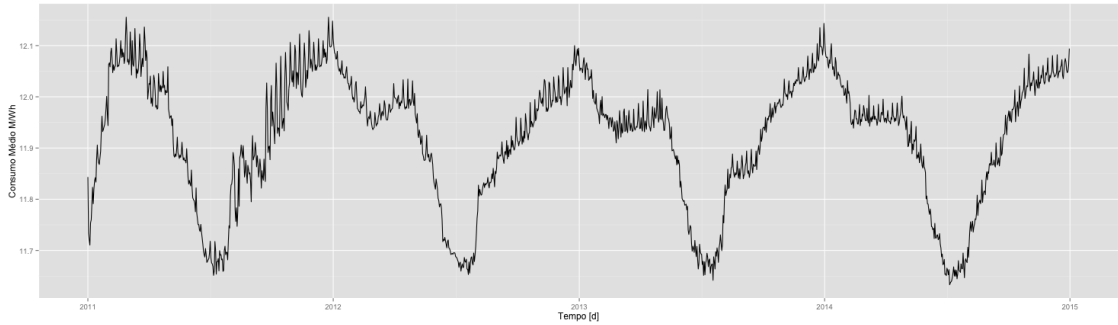


Figura 3.2: Consumo médio (MWh) diário no período de janeiro de 2011 à dezembro de 2014.

A análise da função de autocorrelação permite-nos medir o grau de relação linear entre consumos consecutivos. A Figura 3.3, mostra os gráficos da função de autocorrelação ( $ACF$ ) e função de autocorrelação parcial ( $PACF$ ) da série e revela periodicidade significativa.

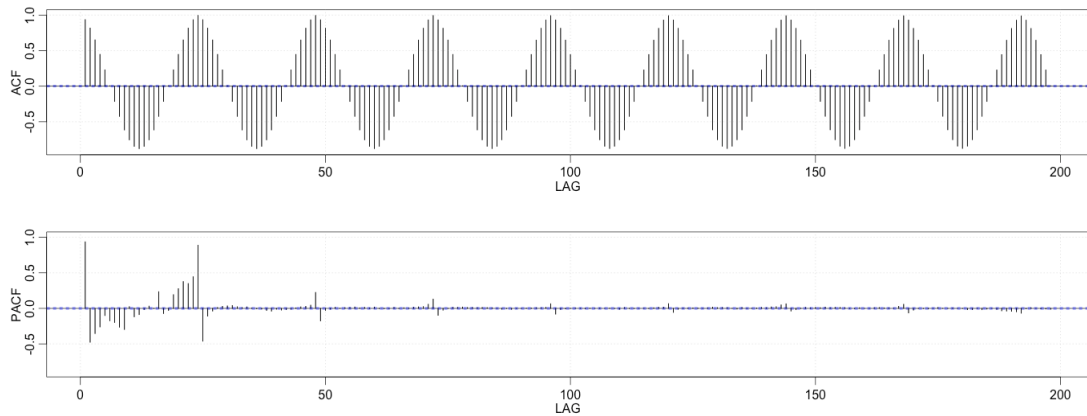


Figura 3.3: Gráficos de  $ACF$  e  $PACF$  da série de energia elétrica com os limites do intervalo de confiança em 95% (tracejado em azul).

O gráfico da  $ACF$  revela um processo autoregressivo sazonal em lento decaimento a partir do  $lag$  0 com um padrão sinusoidal de um comprimento com 24 períodos e todos os  $lags$  estão fora dos limites do intervalo de confiança. Logo, o gráfico  $PACF$ , apresenta uma brusca queda a partir do  $lag$  0 e apenas alguns  $lags$  são significativos (a maioria estão muito próximas dos limites). Na Figura 3.4 é possível observar o consumo (na hora  $t$ ) versus o consumo desfasado em diferentes  $lags$  (hora  $t - i$  e  $i = 1, \dots, 48$ , para a mesma hora do dia  $t - 24$  (no dia anterior) e  $t - 48$  (dois dias antes) e revela que a previsão para as próximas 24 horas é o valor mais auto

correlacionado, enquanto o próximo dia (24 horas) o valor mais auto correlacionado é o que corresponde a uma semana antes.

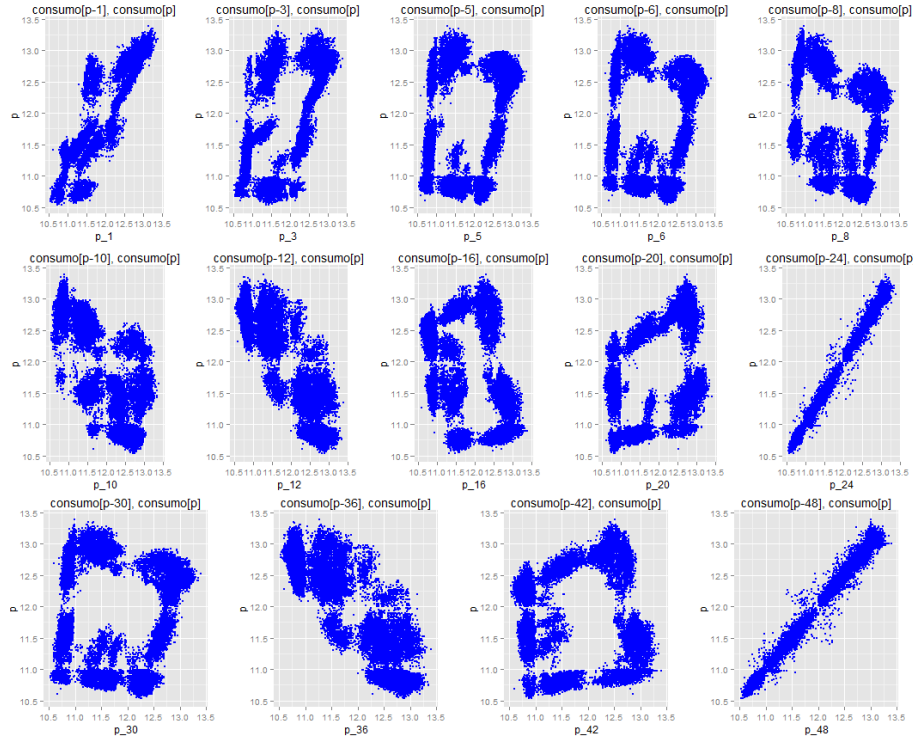


Figura 3.4: Consumo na hora  $t$  versus o consumo desfasado em diferentes  $lags$  ( $t - i$  com  $i = 1, \dots, 48$ ), indica que há uma forte dependência para a mesma hora do dia  $t - 24$  (no dia anterior) e  $t - 48$  (dois dias antes).

## SELEÇÃO E ESTIMAÇÃO DOS MODELOS

O conjunto de dados inicial possui 3 anos de dados (35064 observações) horários. Neste experimento, o  $D_{treino}$  dispõe apenas das observações referentes ao último ano (8760 observações), ou seja, de 01-10-2013 às 00:00 à 30-09-2014 às 23:00. O  $D_{teste}$  é constituído pelas vinte e quatro horas do dia 01-10-2014.

O modelo estatístico selecionado para esse caso foi o ARIMA e a ordem do modelo foi realizada com  $D_{treino}$ , através da função *auto.arima* do pacote *forecast* (Hyndman and Khandakar, 2007) no R Core Team (2014).

A seleção dos preditores para os modelos de aprendizagem automática possui diferentes parâmetros de entrada que dependem principalmente do horizonte da previsão:

- **Vinte e quatro horas à frente**  $H = 24$

Para previsão da série em  $t + k$ , no momento  $k$  o vector de entradas será o



conjunto de  $k$  menos  $\{24, (168 - 24), 168, 169\}$  horas. Como já salientado, o valor mais correlacionado para uma previsão 24 horas à frente é o valor correspondente há uma semana.

Para o modelo MLP (ver Secção 2.2.2), os parâmetros foram estimados por *back-propagation*, o número de camadas ocultas  $NH$  foram variadas dentro do conjunto  $\{5, 7, 9\}$  e combinados as seguintes opções de pesos  $\{0.1, 0.01, 0.001, 0.2, 0.02\}$ , com os resultados obtidos utilizando uma função linear.

Para o modelo BST-UN em cada *base-learner* ( $bbs()$ ) utilizou-se apenas as variáveis explicativas, já os BST-BI o parâmetro  $gl=3$  apresentou melhores resultados que  $gl=4$ . Como justificado em Hofner et al. (2014) a decomposição *P-spline* requer um valor pequeno para  $gl$ , mais detalhes sobre este modelo na Secção 2.2.2.

A seleção desses hiperparâmetros e o modelo final foi dado pelo método de amostragem *cross-validation* com 3 partições.

## ANÁLISE DOS RESULTADOS

A Figura 3.5, apresenta uma comparação entre valores previstos versus observados do consumo de energia elétrica em 24 passos à frente, os resultados estão presentes na Tabela B.1. Na análise da Figura 3.5 é possível considerar que DIR-LIN, RTY-BST-L2 com UN e BI se ajustam razoavelmente a série real observada.

Na Tabela B.2 estão as medidas de precisão RMSE, MAPE e MASE (ver Secção 2.4) que avaliaram o desempenho das diferentes estratégias e modelos ao longo dos  $H = 24$  horizontes de previsão, onde é possível tirar as mesmas conclusões de todas as medidas de precisão, porém apenas o MAPE será utilizado para comparar as estratégias. A evolução dos erros das previsões é apresentado na Figura 3.6 e na Tabela B.3 estão as medidas de previsões globais das diferentes estratégias e modelos no consumo de energia elétrica em Cabinda.

A partir dos resultados observados conclui-se que, para os índices MAPE a estratégia direta linear apresenta os menores valores a partir do 4 passos à frente ( $h$ ), de 8 à 11 e entre 15 a 18 passos à frente. A estratégia de retificação L1-B1 tem boas previsões no 3 e 23 passos à frente, mas não superior a L1-UN que teve melhor desempenhos para 6 e 7, e de 20 à 22 passos à frente. Da comparação e análise das estratégias é possível concluir que:

- Há um desempenho comparável entre os modelos da estratégia recursiva e o modelo ARIMA teve melhor expressão que os demais. REC-LIN e REC-MLP, tiveram um comportamento semelhante entre si, indicando que o modelo MLP seleciona estruturas semelhantes ao modelo LIN. Um desempenho oposto acontece nas estratégias diretas, DIR-MLP tem erros mais elevados do que DIR-LIN entre  $h = 8$  e  $h = 10$  esse comportamento volta aparecer entre  $h = 17$  e  $h = 23$ , indicando que o modelo MLP selecionou estruturas que são bastante diferente ao modelo LIN.



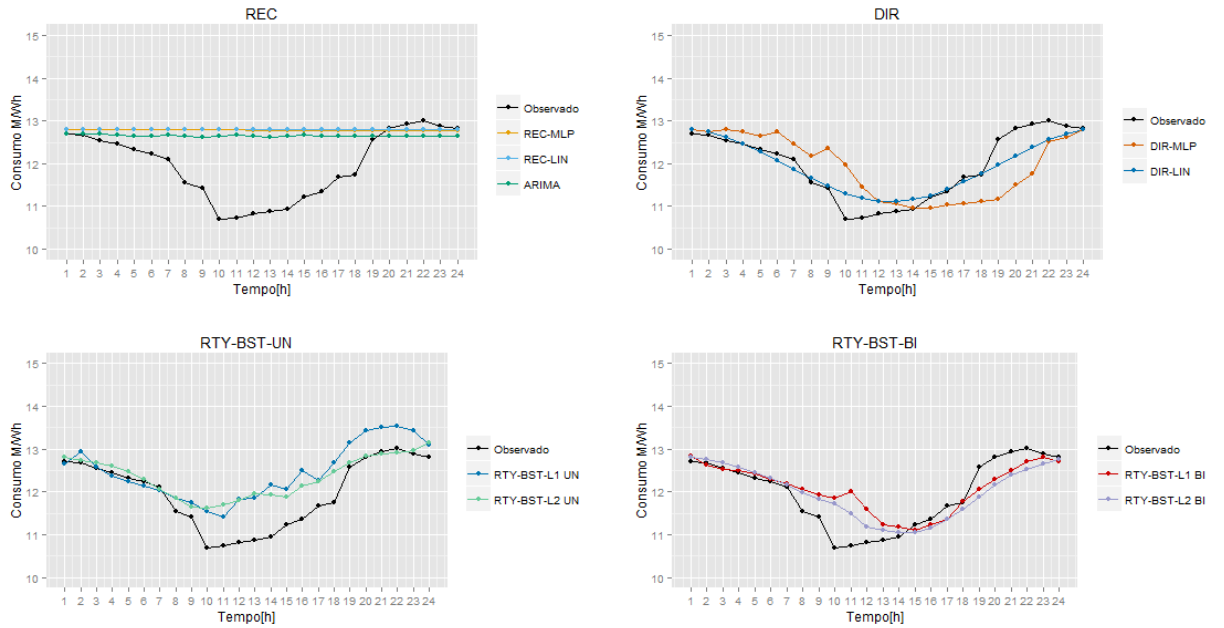


Figura 3.5: Gráficos dos resultados das diferentes estratégias para comparação dos valores previstos versus observados (cor preta) do consumo de energia elétrica em 24 passos à frente.

- As estratégias DIR têm desempenho geral superior às estratégias REC. Uma explicação pode estar no comprimento da série de Cabinda ( $T = 8760$ ), menos ruídos e o longo horizonte de previsão ( $H = 24$ ). [Ben Taieb \(2014\)](#), explica que a estratégia DIR pode fornecer melhores previsões que as REC, uma vez que não sofre com a propagação de erros com os horizontes. Além disso, neste caso,  $T$  é grande em relação ao horizonte  $H$ , pelo que a perda de dados (da estratégia DIR) com o horizonte não tem influência nos resultados.
- As estratégias de RTY-BST com função de perda  $L2$  têm desempenho comparável entre BI e UN e superior quando comparado aos demais modelos de todas as estratégias, onde é evidente a diminuição dos erros nos horizontes mais curtos. Isto confirma a vantagem da estratégia de retificação com BST, modelando os resíduos previsão geradas por REC-LIN usando-o como modelo direto não linear.
- Em particular, a retificação com o BST-BI obteve melhor interpretação que o BST-UN. Quanto a função de perda,  $L2$  tem desempenho superior a  $L1$  em ambos  $P$ -splines. A ideia desta função é penalizar grandes desvios em relação à saída, negligenciando pequenos resíduos.

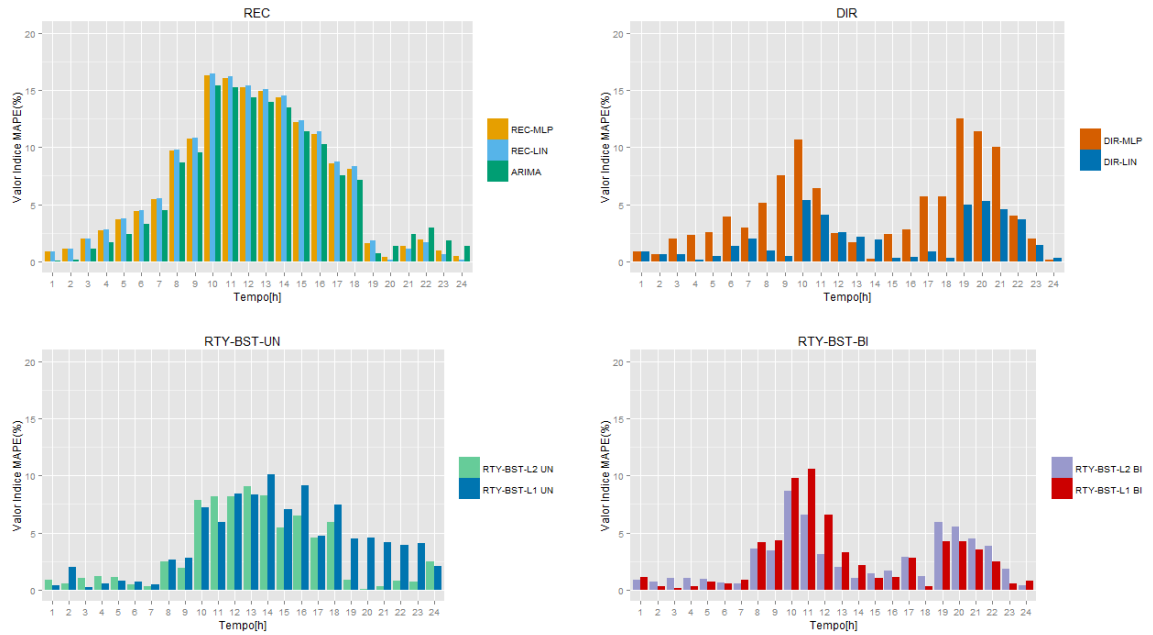


Figura 3.6: Gráficos dos índices MAPE(%) que avaliaram o desempenho das previsões nas diferentes estratégias e modelos no consumo de energia elétrica com  $H = 24$ .

## 3.2 Caso 2 - Retornos do New York Stock Exchange

Criada em 1792 a *New York Stock Exchange* (NYSE) é a bolsa de valores localizada em Manhattan no estado de Nova Iorque nos Estados Unidos da América. Esta é a maior bolsa de valores do mundo, onde são transacionadas ações de grandes empresas. A série temporal em estudo pertence ao mercado financeiro e possui valores diários referentes aos retornos da NYSE de 02-02-1984 à 31-12-1991, estes dados foram extraídos do pacote *astsa* (Shumway, 1988) do R Core Team (2014).

A Figura 3.7, revela uma brusca queda em 19 de outubro de 1987 (Shumway and Stoffer, 2013). O comportamento dos dados apresentados neste gráfico são típicos em séries financeiras, onde a média da série parece ser estável e aproximadamente zero, no entanto, a volatilidade (ou variação) dos dados altera-se ao longo do tempo.

O gráfico da *ACF* na Figura 3.8, revela um brusco decaimento para 0 a partir do *lag* 1 e poucos *lags* são expressivos, ou seja, fora das bandas de confiança. Portanto, é possível concluir que para uma previsão de cinco dias à frente, os valores mais correlacionado são aqueles que corresponde a um e cinco dias antes.

### SELEÇÃO E ESTIMAÇÃO DOS MODELOS

O conjunto de dados inicial possui 2000 observações diárias. Neste caso, o  $D_{treino}$  terá 1945 (incluindo 19-10-1987) observações e  $D_{teste}$ , usa às cinco observações seguintes.

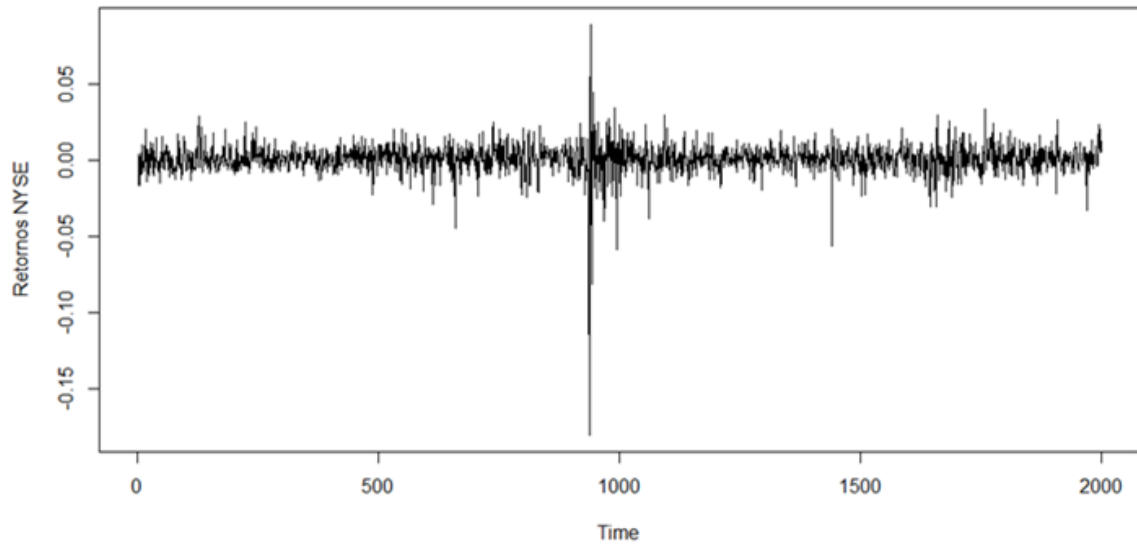


Figura 3.7: Retornos da NYSE. Os dados são valores ponderados dos retornos no mercado, a partir de 02 de fevereiro de 1984 à 31 de dezembro de 1991 (2000 dias de negociação). 19 de outubro de 1987 ocorre em  $t = 938$ .

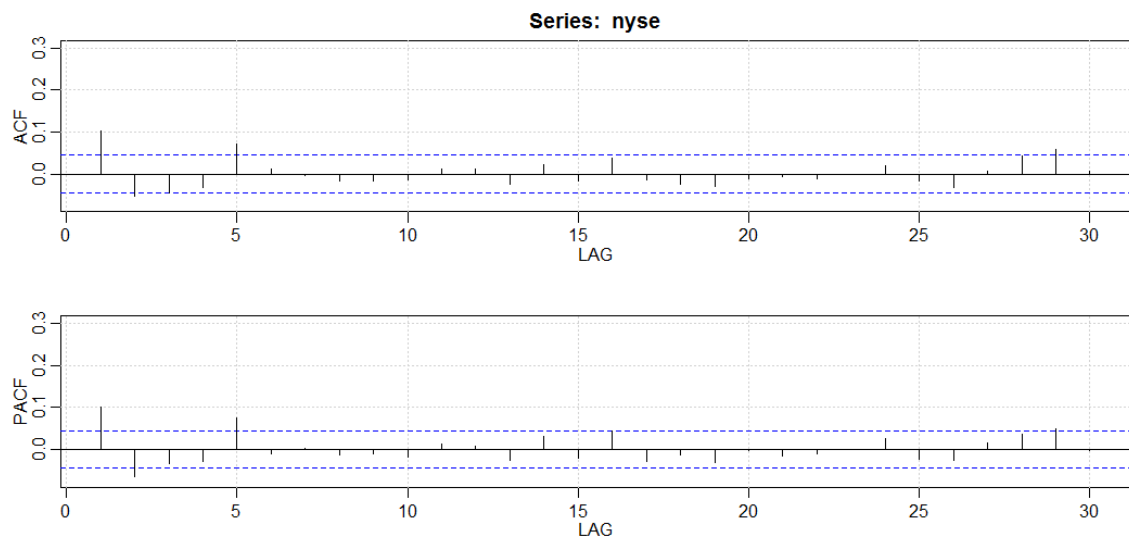


Figura 3.8: Gráficos de  $ACF$  e  $PACF$  dos retornos da NYSE, com os limites do intervalo de confiança em 95% (tracejado em azul).

O modelo estatístico utilizado neste caso foi o GARCH e o parâmetro do modelo é GARCH(1,1), para modelar a série com um AR(1) médio condicional, contudo o pacote *fGarch* do [R Core Team \(2014\)](#) foi utilizado para ajustar o modelo.

Como no caso anterior, a seleção dos preditores para os modelos de aprendizagem automática irá ter diferentes parâmetros de entrada que dependem do horizonte da previsão, segue definido como:

- **Cinco dias à frente  $H = 5$**

Para previsão da série em  $t + k$ , no momento  $k$  o vector de entradas será o conjunto de  $k$  menos  $\{1, 5\}$  dias. Como visto anteriormente, o valor mais correlacionado para uma previsão 5 dias à frente corresponde ao dia e a semana anterior.

Para o modelo MLP (ver Secção 2.2.2), os parâmetros foram estimados por *back-propagation*, o número de camadas ocultas  $NH$  foram variadas dentro do conjunto  $\{1, 2, 3, 4, 5, 6\}$ , combinados com os pesos  $\{0.1, 0.01, 0.001, 0.2, 0.02, 0.002\}$  e com os resultados obtidos utilizando uma função linear.

Assim como em Cabinda os modelos BST-UN e BST-BI, tiveram o mesmo comportamento, onde o parâmetro  $gl=3$  gerou resultados mais expressivos que  $gl=4$ .

O mesmo método foi utilizado para seleção destes hiperparâmetros, o *cross-validation* com 5 partições.

## ANÁLISE DOS RESULTADOS

A Figura 3.9, mostra uma comparação entre valores previstos versus observados dos retornos da NYSE 5 passos à frente, os resultados estão presentes na Tabela B.4. Através da Figura 3.9 não é possível encontrar nenhum modelo que se ajuste à série observada. As características dos sistemas que geram as séries financeiras (em especial a não-linearidade, complexidade e a não-estacionaridade) resultam em vários aspectos que condicionam a eficiência dos vários possíveis modelos de previsão aplicáveis.

Na Tabela B.5, estão as medidas de precisão RMSE e MASE (ver Secção 2.4) que avaliaram o desempenho das diferentes estratégias e modelos ao longo dos  $H = 5$  horizontes de previsão e a evolução dos erros está representado na Figura 3.10. A Tabela B.6, mostra as medidas de previsões globais das diferentes estratégias e modelos. A medida MAPE foi desconsiderada deste caso, pois têm uma desvantagem de ser infinita ou indefinida se  $y_i = 0$  para qualquer  $i$  dentro do período de interesse e tendo valores extremos quando qualquer  $y_i$  está próximo de zero (Hyndman and Athanasopoulos, 2014). Isso é exatamente o que acontece com os resultados dos retornos, estão muito perto de zero. No entanto, apenas a medida RMSE será utilizada para comparar as estratégias e os resultados.

A partir dos resultados observados conclui-se que, para os índices RMSE a estratégia recursiva GARCH apresenta menor valor para 2, 3 e 5 passos à frente, portanto melhor modelo em geral. Para previsão 1 passo à frente os modelos lineares são superiores, mas não supera o direto que teve menores erros para previsão 1 e 4 passos

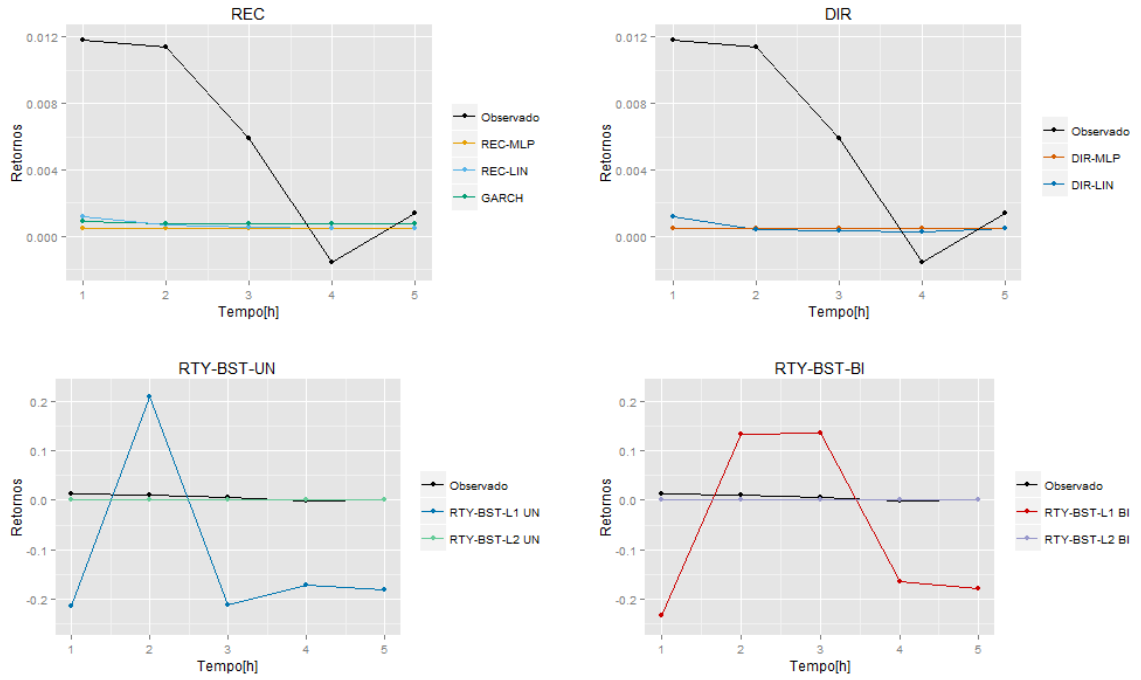


Figura 3.9: Gráficos dos resultados das diferentes estratégias para comparação dos valores previstos versus observados (cor preta) dos retornos da NYSE em 5 passos à frente.

à frente. Ainda é possível observar que os modelos da estratégia RTY-BST utilizando a função de perda  $L1$  não trouxeram nenhuma vantagem para este caso. Da comparação e análise das estratégias é possível concluir que:

- As estratégias REC têm desempenho comparável a DIR em todos os  $h$ . Entre os modelos DIR-LIN e DIR-MLP há um desempenho muito semelhante. O mesmo comportamento pode ser observado quando comparado REC-MLP e REC-LIN. Assim como no caso anterior, o modelo MLP selecionou uma estrutura muito próxima do modelo LIN.
- As estratégias diretas têm desempenho superior a RTY-BST. Isto sugere que os ajustes das previsões recursivas lineares não fornecem qualquer benefício na retificação BST de ambos  $P$ -splines neste caso.
- Nas estratégias de RTY-BST a função de perda  $L2$  têm um comportamento comparável (nos dois  $P$ -splines) a DIR ao longo de todos os horizontes. Mais uma vez, a redução da variância dos componentes *boosting* pode explicar o melhor desempenho para RTY-BST-L2.

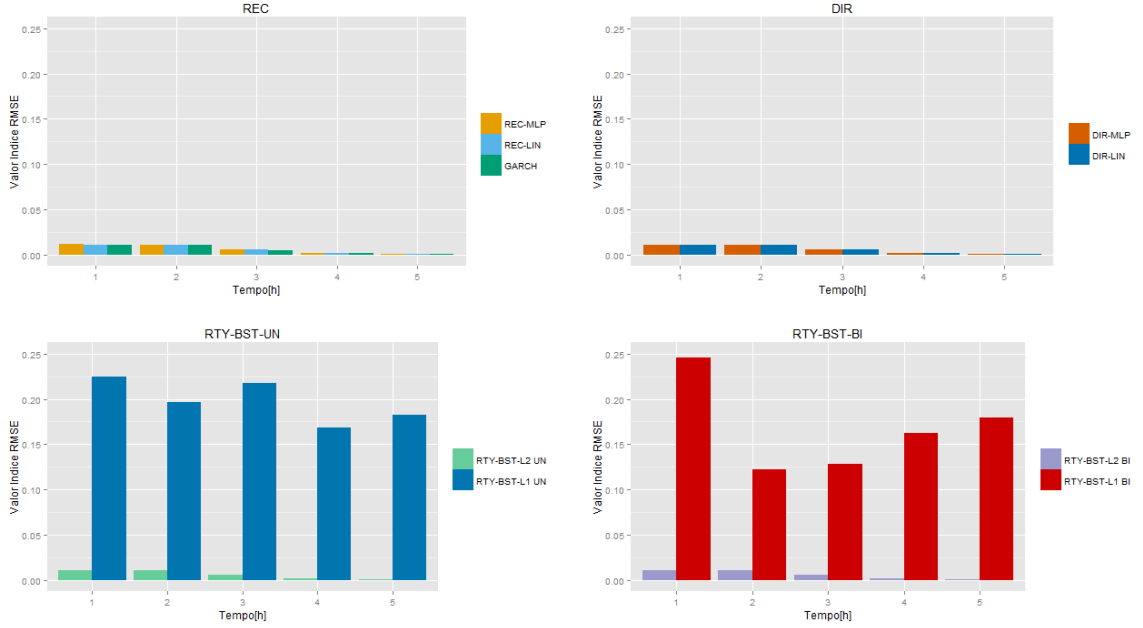


Figura 3.10: Gráficos dos índices RMSE que avaliaram o desempenho das previsões nas diferentes estratégias e modelos dos retornos da NYSE com  $H = 5$ .

### 3.3 Caso 3 - Índice de Produção Mensal do *Federal Reserve Board*

O *Federal Reserve Board* é o principal órgão regulador do Sistema de Reserva Federal encarregado de supervisionar os Bancos da Reserva Federal e auxilia na implementação da política monetária nos Estados Unidos (Wikipedia, 2015). A série temporal desta experiência descreve o índice de produção mensal do *Federal Reserve Board* de 01 janeiro de 1948 à 31 dezembro de 1978 (372 observações mensais), como elucida a Figura 3.11. Esta série também foi extraída do pacote *astsa* (Shumway, 1988) do R.

Na Figura 3.11 é evidente que ocorreu uma queda entre 1974 e 1975 no índice de produção, estabilizou no ano seguinte e durante os últimos anos de estudo sucedeu um aumento. De modo geral os índices aumentam ao longo do tempo. Através dos gráficos de decomposição na Figura 3.12 é possível concluir que a série apresenta uma tendência linear crescente, uma sazonalidade de período 12 e resíduos que aparecem ter ainda estrutura de correlação, principalmente entre 1974 e 1975.

O gráfico da *ACF* na Figura 3.13, revela um lento decaimento para zero e todos os *lags* estão fora dos intervalos de confiança, os quais apresentam certa instabilidade, pelo que é possível concluir que a série não é estacionária. Logo, realizou-se a análise das funções *nsdiffs* (função que estima o número de diferenças simples) e *ndiffs*

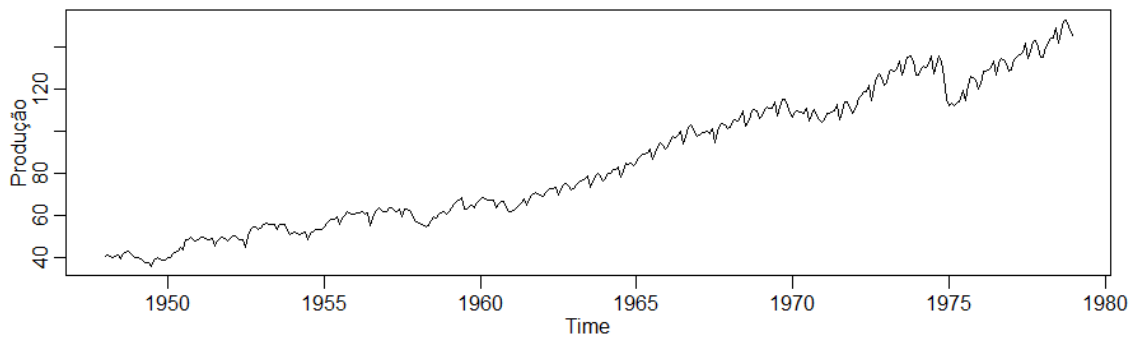


Figura 3.11: Índice de produção mensal do Sistema de Reserva Federal (1948-1978,  $T = 372$  meses).

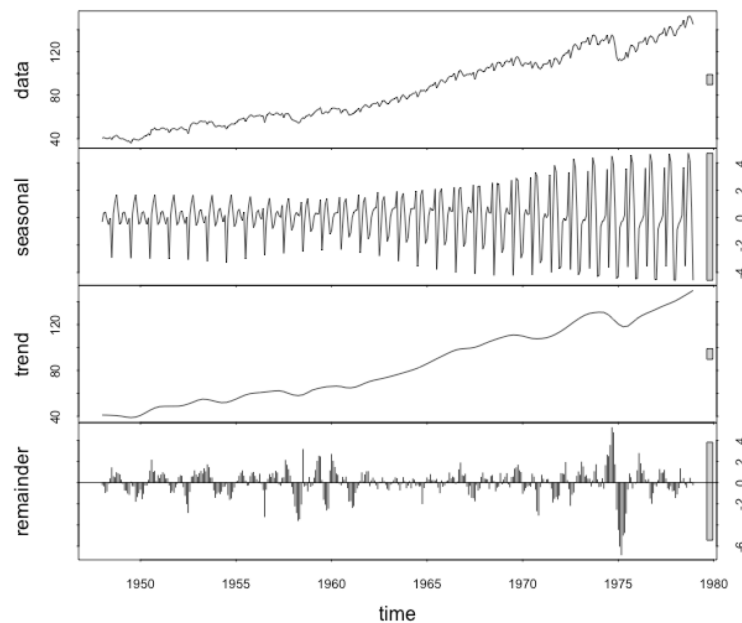


Figura 3.12: Decomposição dos principais componente (sazonal, tendência e ciclo) do índice de produção mensal do Sistema de Reserva Federal (1948-1978).

(estima o número de diferenças sazonais), que resultou numa boa interpretação uma diferenciação simples. Após uma simples diferenciação da série é possível observar que a função de autocorrelação (ver Figura 3.14), decresce bruscamente a partir do *lag 1*, apresentando poucos *lags* fora das bandas de confiança. Entretanto, é possível concluir que os valores mais correlacionados correspondem aos dois primeiros meses, os três últimos e o ano antecedente.

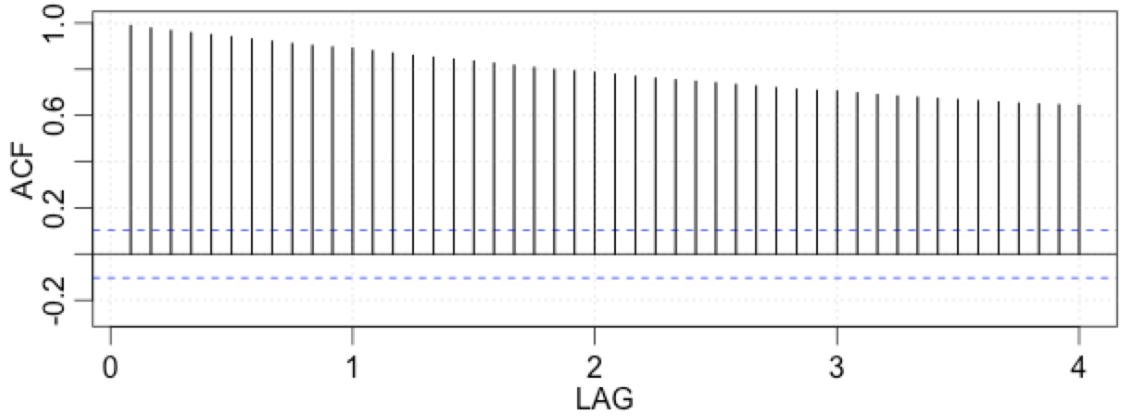


Figura 3.13: Gráfico da  $ACF$  do índice de produção, com os limites do intervalo de confiança em 95% (tracejado em azul) e  $lag$  máximo de 48.

### SELEÇÃO E ESTIMAÇÃO DOS MODELOS

O conjunto de dados inicial possui 3 anos de dados (372 observações) mensais de janeiro de 1948 a dezembro de 1978. Neste estudo, o  $D_{treino}$  usa apenas das observações referentes os últimos dois anos (360 observações), já  $D_{teste}$  possui os valores observados do último ano (12 meses).

Para esse caso foi proposto um modelo ARIMA sazonal e através dos correlogramas (Figura 3.14) a ordem do modelo é definida por SARIMA(1,1,0)(0,1,1).

Assim como nos dois casos anteriores, a seleção dos preditores para os modelos de aprendizagem automática irá ter diferentes parâmetros de entrada que dependem do horizonte da previsão. Assim

- **Doze meses à frente  $H = 12$**  - Para previsão da série em  $t + k$ , no momento  $k$  o vector de entradas será o conjunto de  $k$  menos  $\{1, 2, 3, 4, 9, 10, 11, 12, 24\}$  horas. As entradas 1 e 4 e os demais correspondem aos valores mais correlacionado para uma previsão 12 meses à frente.

Para o modelo MLP (ver Secção 2.2.2), os parâmetros foram estimados por *back-propagation*, o número de camadas ocultas  $NH$  foram variadas dentro do conjunto  $\{4, 5, 7, 8\}$  e combinados as seguintes opções de pesos  $\{0.1, 0.01, 0.2, 0.02\}$  com os resultados obtidos utilizando uma função linear.

Assim como em Cabinda e NYSE os modelos BST-UN e BST-BI, tiveram o mesmo comportamento e o método utilizado para seleção dos hiperparâmetros foi o *cross-validation* com 4 partições.



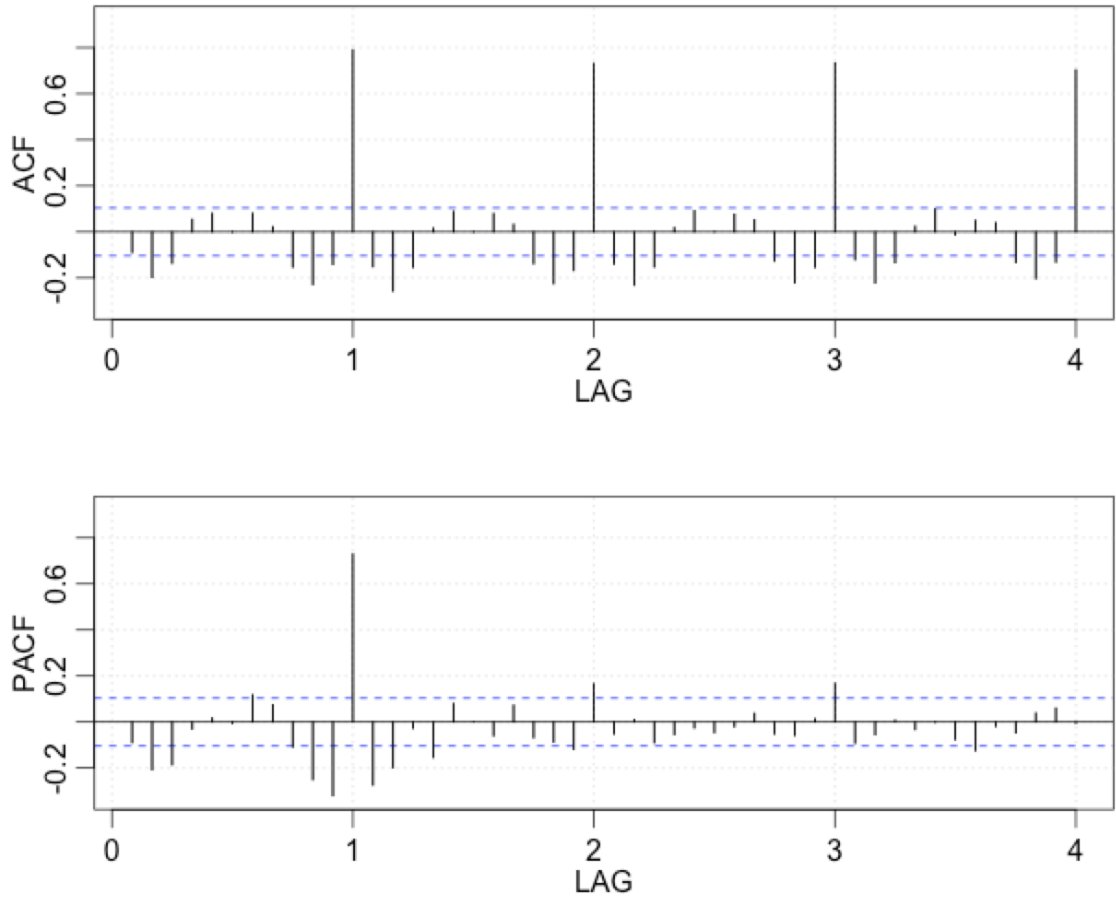


Figura 3.14: Gráficos de  $ACF$  e  $PACF$  do índice de produção com uma diferença simples e os limites do intervalo de confiança em 95% (tracejado em azul) e  $lag$  máximo de 48.

### ANÁLISE DOS RESULTADOS

A Figura 3.15, mostra uma comparação entre valores previstos versus observados do índice de produção mensal do Sistema de Reserva Federal 12 passos à frente, os resultados estão presentes na Tabela B.7. Através da Figura 3.15 a estratégia recursiva com o modelo ARIMA sazonal teve um bom ajuste em relação a série real observada.

Na Tabela B.8 estão as medidas de precisão RMSE, MAPE e MASE (ver Secção 2.4) que avaliaram o desempenho das diferentes estratégias e modelos ao longo dos ( $H = 12$ ) horizontes de previsão e é possível tirar as mesmas conclusões de todas as medidas de precisão, porém apenas o MAPE será utilizado para comparar as estratégias. A evolução dos erros está representado na Figura 3.16 e na Tabela B.7,

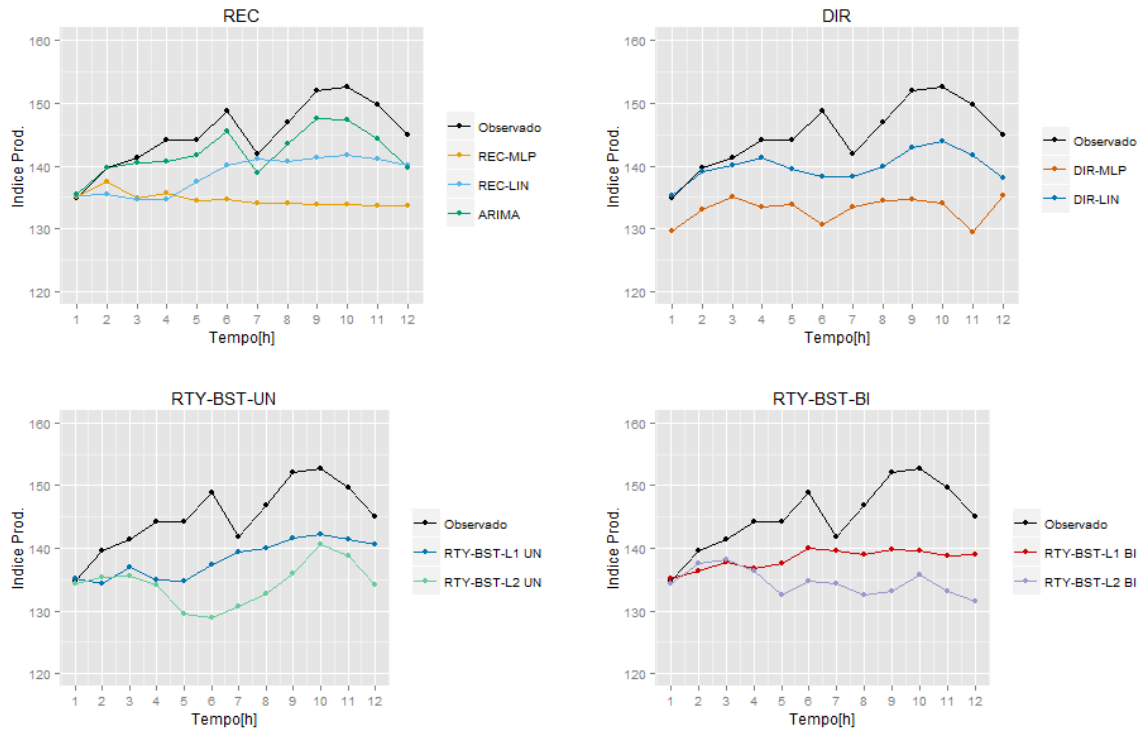


Figura 3.15: Gráficos dos resultados das diferentes estratégias e modelos para comparação dos valores previstos versus observados do índice de produção mensal do Sistema de Reserva Federal 12 passos à frente.

estão as medidas de previsões globais das diferentes estratégias e modelos no índice de produção.

A partir dos resultados observados conclui-se que, para os índices MAPE o modelo não linear recursivo teve bom resultado apenas no primeiro passo à frente. O recursivo linear ARIMA têm menor erro para as previsões 2 e 3 passos à frente, 5 e 6 e da previsão 8 a 11 passos à frente, considerado o melhor modelo na globalidade. A estratégia de retificação UN-L2 tem boa representação apenas para a previsão 12 passos à frente. É ainda evidente os modelos MLP (REC e DIR) geraram previsões mais pobres para este caso. Da comparação e análise das estratégias é possível concluir que:

- A estratégia DIR têm desempenho comparável com a estratégia recursiva modelo ARIMA sazonal apenas nos horizontes mais curtos ( $h = 1, 2, 3$ ), mas tem erros mais elevados nos horizonte mais distantes ( $h = 10, 11, 12$ ).
- As estratégias REC-MLP apontam erros mais elevados que REC-LIN. O mesmo comportamento é observado entre as estratégias DIR-MLP e DIR-LIN. Di-

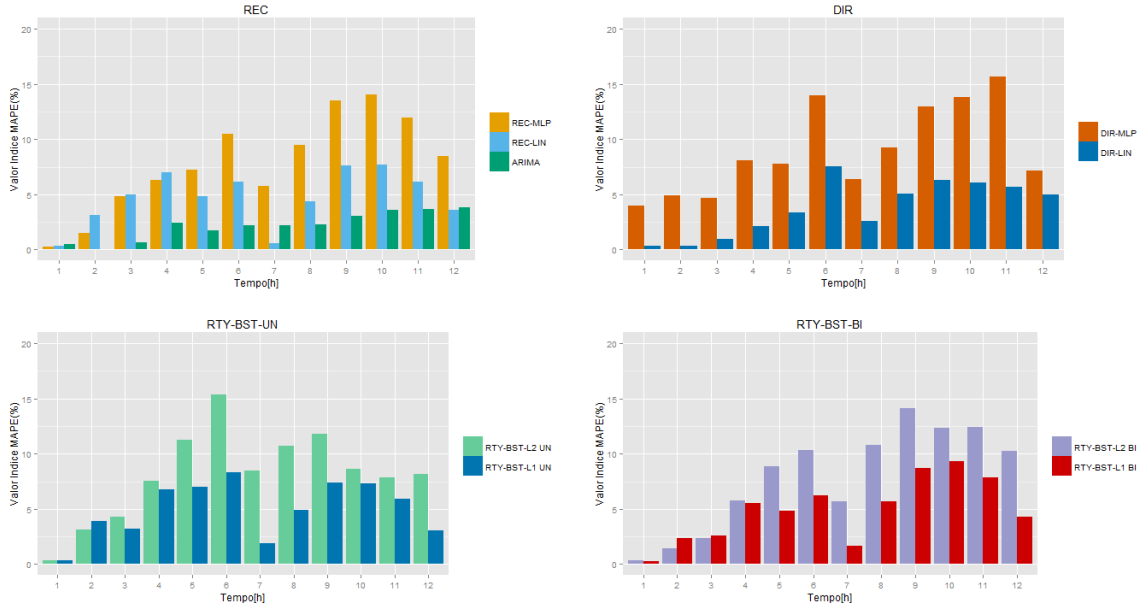


Figura 3.16: Gráficos dos índice MAPE (%) que avaliaram o desempenho das previsões nas diferentes estratégias e modelos do índice de produção mensal do Sistema de Reserva Federal com  $H = 12$ .

ferente do que acontece na série NYSE (Secção 3.2), o modelo MLP pode selecionar muitas vezes uma estrutura diferente dos modelos LIN.

- Ao contrário do que acontece em Cabinda (Secção 3.1), as estratégias REC têm desempenho ligeiramente melhor que DIR. Uma possível razão é que a estratégia DIR passa por uma diminuição da amostra com os horizontes de previsão, o que pode ser importante neste caso para os horizontes mais distantes ( $H = 12$ ) os dados utilizados no treino podem ser relativamente curtos ( $T < 360$ ).
- Os modelos da estratégia RTY têm desempenho comparável apenas nos primeiros horizontes ( $h = 1, 2, 3$ ). Ao contrário do que acontece na série de Cabinda (Secção 3.1) a função de perda  $L2$  apresenta elevados erros em ambos  $P$ -splines, isto sugere que o modelo BST com esta função de perda não apresentou benefício nenhum quando comparado com  $L1$ . Logo, a estratégia RTY-BST com  $L1$  aponta desempenho superior estratégia DIR-MLP, sugerindo que novamente a variância foi reduzida com os componentes BST.

### 3.4 Considerações Finais

Em todos os casos foram obtidos resultados diferentes. De fato, as estratégias recursivas superaram a direta em todos os modelos (algoritmos de aprendizagem e estatística) nos dois últimos casos e vice-versa no primeiro caso. Em particular, as previsões recursivas lineares fornecem melhores resultados para os Casos 2 e 3, mas são superado por outros modelos no Caso 1. Consequentemente, as estratégias de retificação com *boosting* geram boas previsões para o Caso 1, mas fornecem previsões mais pobres nos Caso 2 e 3 quando comparado com as previsões recursivas e diretas. No entanto, a seleção entre os *base-learner* (UN e BI) e a função de perda do algoritmo *boosting* são decisivas. No Caso 2,  $L2$  supera  $L1$  e vice-versa para o Caso 3. Na verdade, a ideia da estratégia de retificação é que as previsões recursivas lineares forneçam uma boa primeira aproximação que necessitam apenas de alguns ajustes. Esta suposição parece ser confirmada somente na série de cabinda (Caso 1).

## Capítulo 4

### Conclusões

Obter modelos capazes de prever a evolução de sistemas multi-passos à frente é um dos grandes desafios quando se lida com previsão em séries temporais. A busca por modelos que possam extrair da melhor forma as informações existentes nos dados é de grande importância em diversas áreas do conhecimento. A maioria dos estudos centraram-se na comparação entre as previsões recursivas e diretas, investigando as condições em que uma estratégia é melhor do que a outra. Porém, escolher entre essas duas estratégias é uma tarefa desafiadora em aplicações do mundo real, uma vez que encontrar a melhor envolve um *trade-off* entre viés e variância das previsões. Visando preencher essa lacuna foi introduzida, recentemente na literatura uma estratégia de retificação usando o algoritmo gradiente *boosting* que envolve os chamados *weak learner* também conhecido como *base-learner*. Os autores da estratégia de retificação concluem que em séries do mundo real as previsões recursivas lineares geram bons resultados precisando apenas de alguns ajustes (retificação) em cada horizonte para captar a não linearidade presente em alguns casos. A estratégia de retificação com *boosting* mostrou-se particularmente útil para problemas de previsão multi-passos à frente pela atenção que o algoritmo tem recebido na literatura de previsão.

Neste contexto, o objetivo deste trabalho foi analisar diferentes modelos lineares e não lineares e realizar previsões  $k$ -passos à frente com três estratégias: recursiva, direta e retificação com *boosting*, em três séries temporais com diferentes comportamento, tamanho e horizontes de previsão, comparando-se o desempenho das mesmas segundo uma determinada métrica. Os resultados alcançados estão expostos no Capítulo 3 e mostram que não há uma única estratégia que forneça as melhores previsões para todos os horizontes e em comparação com as estratégias recursiva e diretas, é possível concluir que a estratégia de retificação oferece um desempenho bem próximo ou até superior que a melhor entre as estratégias recursivas e diretas.

No geral, a precisão das previsões multi-passos à frente com base em algoritmos de aprendizado automática podem ser significativamente melhoradas se uma estratégia previsão adequada for utilizada para selecionar os parâmetros do modelo e

para gerar as previsões. O mesmo acontece com os modelos gerados pelos métodos estatísticos. Em suma, a escolha de diferentes modelos aliados a diferentes estratégias multi-passos à frente é uma alternativa viável e que pode levar a melhores resultados no índice de desempenho escolhido no problema de previsão em séries temporais.

## 4.1 Limitações e Trabalho Futuro

O foco deste trabalho é exclusivamente voltado para séries temporais univariadas enquanto na prática, muitas vezes encontramos problemas de previsões multivariadas, onde requer a previsão de um conjunto de séries temporais possivelmente dependente. Uma direção futura importante seria estender essas estratégias para a comportar séries multivariadas.

Os resultados obtidos neste trabalho indicam que vale a pena considerar diferentes as estratégias com diferentes modelos. Dando continuidade a essa linha de raciocínio, sugerem-se alguns aspectos que ainda podem ser explorados:

- Utilizar outros modelos gerados por algoritmos de aprendizagem automática, como SVM (do inglês, *support vector machine*) e compará-lo com o algoritmo de rede neuronal aplicado neste trabalho.
- Aprofundar os estudos nos modelos obtidos pela algoritmo de rede neuronal, por exemplo, acrescentar outro tipo variáveis explicativas e observar se os resultados são aceitáveis.
- Explorar os diferentes componentes do algoritmo *boosting*, como por exemplo, outros *base-learners*.
- Explorar e implementar a previsão multi-passos da estratégia direta com os modelos estatísticos tradicionais lineares e não lineares.

*Cross-validation* é o método mais popular para selecionar os hiperparâmetros dos modelos de aprendizagem automática. No entanto, o *cross-validation* é conhecido por ser inadequado com dados de séries temporais devido à dependência do tempo entre as observações. Há poucos trabalhos, em particular quase não existem na área da pesquisa métodos de *cross-validation* para previsão multi-passo em séries temporais.

# Bibliografia

- Abu-Mostafa, Y. S., Magdon-Ismail, M., and Lin, H.-T. (2012). *Learning from data*. AMLBook.
- Ahmed, N. K., Atiya, A. F., Gayar, N. E., and El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6):594–621.
- Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the institute of Statistical Mathematics*, 21(1):243–247.
- Alpaydin, E. (2004). *Introduction to machine learning*. MIT press.
- Ben Taieb, S. (2014). *Machine learning strategies for multi-step-ahead time series forecasting*. PhD thesis, Université Libre de Bruxelles, Belgium, <http://theses.ulb.ac.be/ETD-db/collection/available/ULBetd-10072014-002434/>.
- Ben Taieb, S., Bontempi, G., Atiya, A. F., and Sorjamaa, A. (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition. *Expert systems with applications*, 39(8):7067–7083.
- Ben Taieb, S. and Hyndman, R. (2014). Boosting multi-step autoregressive forecasts. In *Proceedings of The 31st International Conference on Machine Learning*, pages 109–117.
- Ben Taieb, S. and Hyndman, R. J. (2012). Recursive and direct multi-step forecasting: the best of both worlds. Technical report, Monash University, Department of Econometrics and Business Statistics.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.
- Box, G. E., Jenkins, G. M., and Reinsel, G. C. (2013). *Time series analysis: forecasting and control*. John Wiley & Sons.

- Bühlmann, P. and van de Geer, S. (2011). Boosting and greedy algorithms. In *Statistics for High-Dimensional Data*, pages 387–431. Springer.
- Bühlmann, P. and Yu, B. (2003). Boosting with the  $l_2$  loss: regression and classification. *Journal of the American Statistical Association*, 98(462):324–339.
- Cheng, H., Tan, P.-N., Gao, J., and Scripps, J. (2006). Multistep-ahead time series prediction. In *Advances in Knowledge Discovery and Data Mining*, pages 765–774. Springer.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–73.
- De Gooijer, J. G. and Hyndman, R. J. (2006). 25 years of time series forecasting. *International journal of forecasting*, 22(3):443–473.
- De Livera, A. M., Hyndman, R. J., and Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527.
- Drucker, H. (1997). Improving regressors using boosting techniques. In *ICML*, volume 97, pages 107–115.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.
- Fenske, N., Kneib, T., Hothorn, T., et al. (2011). Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. *Journal of the American Statistical Association*, 106(494):494–510.
- Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139.
- Freund, Y., Schapire, R. E., et al. (1996). Experiments with a new boosting algorithm. In *ICML*, volume 96, pages 148–156.
- Friedman, J., Hastie, T., Tibshirani, R., et al. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, 28(2):337–407.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.



- Gama, C., Carvalho, A., Oliveira, M., Faceli, K., and Lorena, A. (2012). *Extração de Conhecimento de Dados, Data Mining*. Sílabo.
- Hamzaçebi, C., Akay, D., and Kutay, F. (2009). Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. *Expert Systems with Applications*, 36(2):3839–3844.
- Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85.
- Hofner, B., Mayr, A., Robinzonov, N., and Schmid, M. (2014). Model-based boosting in r: a hands-on tutorial using the r package mboost. *Computational Statistics*, 29(1-2):3–35.
- Hyndman, R. J. (2014). Measuring forecast accuracy. *Citeseer*.
- Hyndman, R. J. and Athanasopoulos, G. (2014). *Forecasting: principles and practice*. OTexts.
- Hyndman, R. J. and Khandakar, Y. (2007). Automatic time series for forecasting: the forecast package for r. Technical report, Monash University, Department of Econometrics and Business Statistics.
- Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688.
- Kline, D. M. (2004). Methods for multi-step time series forecasting with neural networks. *Neural Networks in Business Forecasting*, pages 226–250.
- Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145.
- Kuhn, M. (2008). Building predictive models in r using the caret package. *Journal of Statistical Software*, 28(5):1–26.
- Lapedes, A. and Farber, R. (1987). Nonlinear signal processing using neural networks. *Technical Report LA-UR-87-2662*.
- Ljung, L. (1987). System identification: Theory for the user. *PTR Prentice Hall Information and System Sciences Series*, 198.
- Ming, W., Bao, Y., Hu, Z., and Xiong, T. (2014). Multistep-ahead air passengers traffic prediction with hybrid arima-svms models. *The Scientific World Journal*, 2014.

- Natekin, A. and Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, 7.
- Palit, A. K. and Popovic, D. (2006). *Computational intelligence in time series forecasting: theory and engineering applications*. Springer-Verlag.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Sarle, W. S. (1995). Stopped training and other remedies for overfitting. In *Proc. of the 27th symposium on the interface of computing science and statistics*, pages 352–360.
- Schapire, R. E. (1990). The strength of weak learnability. *Machine learning*, 5(2):197–227.
- Schmid, M. and Hothorn, T. (2008). Boosting additive models using component-wise p-splines. *Computational Statistics & Data Analysis*, 53(2):298–311.
- Schwarz, G. (1978). Estimating the dimension of a model the annals of statistics 6 (2), 461–464. URL: <http://dx.doi.org/10.1214/aos/1176344136>.
- Shumway, R. H. (1988). Applied statistical time series analysis. *Prentice Hall Series in Statistics, Englewood Cliffs, NJ: Prentice-Hall, c1988*, 1.
- Shumway, R. H. and Stoffer, D. S. (2013). *Time series analysis and its applications With R Examples*. Springer. Third Edition - Updated.
- Solomatine, D. P. and Shrestha, D. L. (2004). Adaboost. rt: a boosting algorithm for regression problems. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 2, pages 1163–1168. IEEE.
- Sorjamaa, A., Hao, J., Reyhani, N., Ji, Y., and Lendasse, A. (2007). Methodology for long-term prediction of time series. *Neurocomputing*, 70(16):2861–2869.
- Sorjamaa, A. and Lendasse, A. (2006). Time series prediction using dirrec strategy. *European Symposium on Artificial Neural Networks*, pages 143–148.
- Taieb, S. B. and Hyndman, R. (2014). Boosting multi-step autoregressive forecasts. Technical report, Monash University, Department of Econometrics and Business Statistics.
- Tiao, G. C. and Tsay, R. S. (1994). Some advances in non-linear and adaptive modelling in time-series. *Journal of forecasting*, 13(2):109–131.

- Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 245–292.
- Weigend, A. S. (1994). Time series prediction: forecasting the future and understanding the past. *Santa Fe Institute Studies in the Sciences of Complexity*.
- Werbos, P. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences. *Harvard University, Cambridge, MA*.
- Werbos, P. J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, 1(4):339–356.
- Wikipedia, t. f. e. (2015). Federal reserve board of governors. Acessado 19-09-2015.

## Anexo A

# Estratégias de Previsão Multi-passos

### A.1 Estratégia Recursiva

Um exemplo da estratégia recursiva para 3 passos à frente pode ser exemplificada pelas equações:

$$\begin{aligned}\hat{y}(t+1) &= f(y(t), y(t-1), y(t-2), y(t-3)) \\ \hat{y}(t+2) &= f(\hat{y}(t+1), y(t), y(t-1), y(t-2)) \\ \hat{y}(t+3) &= f(\hat{y}(t+2), \hat{y}(t+1), y(t), y(t-1))\end{aligned}\tag{A.1}$$

Para melhor visualização desta estratégia, um exemplo das previsões de 8 passos à frente é fornecido: Dados o modelo

$$\hat{y}(k) = a_1 y(k-1) + a_2 y(k-2) + a_3 y(k-3)\tag{A.2}$$

e o vetor de entrada  $y$  sendo,

$$y = [y(0)y(1) \quad y(2)y(3) \quad y(4)y(5) \quad y(6)y(7) \quad y(8)y(9) \quad y(10)]\tag{A.3}$$

Pode-se então, estimar 8 passos à frente da seguinte maneira:

$$\begin{bmatrix} \hat{y}(3) \\ \hat{y}(4) \\ \hat{y}(5) \\ \hat{y}(6) \\ \hat{y}(7) \\ \hat{y}(8) \\ \hat{y}(9) \\ \hat{y}(10) \end{bmatrix} = \begin{bmatrix} y(2) & y(1) & y(0) \\ \hat{y}(3) & y(2) & y(1) \\ \hat{y}(4) & \hat{y}(3) & y(2) \\ \hat{y}(5) & \hat{y}(4) & \hat{y}(3) \\ \hat{y}(6) & \hat{y}(5) & \hat{y}(4) \\ \hat{y}(7) & \hat{y}(6) & \hat{y}(5) \\ \hat{y}(8) & \hat{y}(7) & \hat{y}(6) \\ \hat{y}(9) & \hat{y}(8) & \hat{y}(7) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

## A.2 Estratégia Direta

Um exemplo da estratégia direta para 3 passos à frente pode ser exemplificada pelas equações:

$$\begin{aligned}\hat{y}(t+1) &= f_1(y(t), y(t-1), y(t-2), y(t-3)) \\ \hat{y}(t+2) &= f_2(y(t), y(t-1), y(t-2), y(t-3)) \\ \hat{y}(t+3) &= f_3(y(t), y(t-1), y(t-2), y(t-3))\end{aligned}\tag{A.4}$$

Para melhor visualização desta estratégia, um exemplo das previsões de 3 passos à frente é fornecido:

Considerando o vetor  $y$  de entrada sendo,

$$y = [y(0)y(1) \quad y(2)y(3) \quad y(4)y(5) \quad y(6)y(7) \quad y(8)y(9) \quad y(10)]\tag{A.5}$$

1. Previsão de 1 passo à frente:

$$y(k) = a_1y(k-1) + a_2y(k-2) + a_3y(k-3)\tag{A.6}$$

$$\begin{bmatrix} \hat{y}(3) \\ \hat{y}(4) \\ \hat{y}(5) \\ \hat{y}(6) \\ \hat{y}(7) \\ \hat{y}(8) \\ \hat{y}(9) \\ \hat{y}(10) \end{bmatrix} = \begin{bmatrix} y(2) & y(1) & y(0) \\ y(3) & y(2) & y(1) \\ y(4) & y(3) & y(2) \\ y(5) & y(4) & y(3) \\ y(6) & y(5) & y(4) \\ y(7) & y(6) & y(5) \\ y(8) & y(7) & y(6) \\ y(9) & y(8) & y(7) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

2. Previsão de 2 passos à frente:

$$y(k) = b_1y(k-1) + b_2y(k-2) + b_3y(k-3)\tag{A.7}$$

$$\begin{bmatrix} \hat{y}(4) \\ \hat{y}(5) \\ \hat{y}(6) \\ \hat{y}(7) \\ \hat{y}(8) \\ \hat{y}(9) \\ \hat{y}(10) \end{bmatrix} = \begin{bmatrix} y(2) & y(1) & y(0) \\ y(3) & y(2) & y(1) \\ y(4) & y(3) & y(2) \\ y(5) & y(4) & y(3) \\ y(6) & y(5) & y(4) \\ y(7) & y(6) & y(5) \\ y(8) & y(7) & y(6) \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

3. Previsão de 3 passos à frente:

$$y(k) = c_1 y(k-1) + c_2 y(k-2) + c_3 y(k-3) \quad (\text{A.8})$$

$$\begin{bmatrix} \hat{y}(5) \\ \hat{y}(6) \\ \hat{y}(7) \\ \hat{y}(8) \\ \hat{y}(9) \\ \hat{y}(10) \end{bmatrix} = \begin{bmatrix} y(2) & y(1) & y(0) \\ y(3) & y(2) & y(1) \\ y(4) & y(3) & y(2) \\ y(5) & y(4) & y(3) \\ y(6) & y(5) & y(4) \\ y(7) & y(6) & y(5) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

## Anexo B

### Estudo de Casos

#### B.1 Resultados das Previsões Multi-Passos

Tabela B.1: Valor real observado com os valores previstos do consumo de energia elétrica ao longo dos horizontes de previsões vinte quatro passos à frente  $H = 24$  em Cabinda

Estratégias	Horizonte de Previsão (h)																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
<b>REAL</b>	<b>12.7</b>	<b>12.67</b>	<b>12.55</b>	<b>12.45</b>	<b>12.33</b>	<b>12.23</b>	<b>12.1</b>	<b>11.55</b>	<b>11.42</b>	<b>10.7</b>	<b>10.73</b>	<b>10.83</b>	<b>10.87</b>	<b>10.94</b>	<b>11.22</b>	<b>11.35</b>	<b>11.68</b>	<b>11.74</b>	<b>12.57</b>	<b>12.82</b>	<b>12.94</b>	<b>13.01</b>	<b>12.88</b>	<b>12.82</b>
REC-MLP	12.80	12.80	12.80	12.80	12.79	12.79	12.79	12.79	12.78	12.78	12.78	12.78	12.78	12.77	12.77	12.77	12.77	12.77	12.77	12.76	12.76	12.76	12.76	12.76
REC-LIN	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80	12.80
DIR-MLP	12.80	12.74	12.80	12.74	12.64	12.73	12.46	12.16	12.35	11.97	11.46	11.10	11.05	10.96	10.95	11.04	11.05	11.11	11.17	11.51	11.76	12.51	12.62	12.80
ARIMA	12.70	12.69	12.69	12.65	12.63	12.64	12.66	12.64	12.62	12.64	12.66	12.64	12.62	12.64	12.65	12.64	12.63	12.64	12.65	12.64	12.63	12.64	12.65	12.64
RTY-BST-L1 BI	12.84	12.62	12.52	12.49	12.41	12.30	12.20	12.05	11.93	11.86	11.99	11.59	11.24	11.17	11.10	11.23	11.36	11.77	12.06	12.29	12.50	12.69	12.81	12.71
RTY-BST-L2 BI	12.80	12.76	12.68	12.57	12.44	12.31	12.16	11.98	11.82	11.71	11.48	11.18	11.09	11.05	11.06	11.16	11.35	11.59	11.86	12.15	12.38	12.53	12.64	12.76
RTY-BST-L1 UN	12.65	12.92	12.58	12.38	12.23	12.14	12.03	11.86	11.74	11.53	11.40	11.82	11.86	12.17	12.06	12.49	12.26	12.68	13.15	13.42	13.50	13.53	13.42	13.09
RTY-BST-L2 UN	12.81	12.74	12.67	12.59	12.47	12.28	12.06	11.84	11.64	11.61	11.68	11.79	11.94	11.92	11.86	12.13	12.23	12.48	12.67	12.82	12.89	12.91	12.96	13.14



Tabela B.2: Medidas de precisão ao longo de diferentes horizontes para previsões vinte e quatro passos à frente no consumo de energia elétrica em Cabinda.

Estratégias	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
<b>RMSE</b>																								
REC-MLP	0.1067	0.1370	0.2521	0.3502	0.4681	0.5657	0.6935	1.2414	1.3693	2.0872	2.0552	1.9533	1.9114	1.8395	1.5576	1.4258	1.0941	1.0324	0.2007	0.0509	0.1725	0.2440	0.1156	0.0571
REC-LIN	0.1084	0.1389	0.2569	0.3569	0.4767	0.5763	0.7059	1.2556	1.3853	2.1049	2.0746	1.9742	1.9339	1.8636	1.5832	1.4529	1.1226	1.0622	0.2319	0.0184	0.1387	0.2090	0.0794	0.0197
DIR-MLP	0.1069	0.1376	0.2528	0.3512	0.4697	0.5672	0.6967	1.2414	1.3693	2.0872	2.0552	1.9533	1.9114	1.8395	1.5576	1.4258	1.0941	1.0324	0.2007	0.0509	0.1725	0.2440	0.1156	0.0571
DIR-LIN	0.1084	0.1389	0.2569	0.3569	0.4767	0.5763	0.7059	1.2556	1.3853	2.1049	2.0746	1.9742	1.9339	1.8636	1.5832	1.4529	1.1226	1.0622	0.2319	0.0184	0.1387	0.2090	0.0794	0.0197
ARIMA	<b>0.0049</b>	<b>0.0230</b>	0.1451	0.2098	0.3007	0.4175	0.5662	1.0965	1.2062	1.9423	1.9304	1.8156	1.7583	1.7013	1.4374	1.2958	0.9508	0.9010	<b>0.0850</b>	0.1739	0.3070	0.4565	0.1857	0.0337
RTY-BST-L2 BI	0.1084	0.0944	0.1328	0.1279	0.1159	0.0800	0.0677	0.4331	0.4090	1.0134	0.7551	0.3504	0.2233	0.1110	0.1569	0.1873	0.3271	0.1412	0.7007	0.6667	0.5556	0.4773	0.2322	0.0524
RTY-BST-L1 BI	0.1425	0.0406	<b>0.0204</b>	0.0409	0.0867	0.0711	0.1011	0.5011	0.5127	1.1627	1.2684	0.7633	0.3703	0.2393	0.1176	0.1197	0.3129	0.0397	0.5067	<b>0.5209</b>	0.4359	0.3154	<b>0.0670</b>	0.1013
RTY-BST-L2 UN	0.1134	0.0705	0.1264	0.1472	0.1427	<b>0.0531</b>	<b>0.0357</b>	0.2925	0.2223	0.9132	0.9572	0.9656	1.0785	0.9851	0.6478	0.7856	0.5585	0.7414	0.1066	<b>0.0040</b>	<b>0.0424</b>	<b>0.0972</b>	0.0861	0.3226
RTY-BST-L1 UN	0.0478	0.2582	0.0316	0.0699	0.0951	0.0880	0.0603	0.3119	0.3276	0.8330	0.6791	0.9941	0.9924	1.2344	0.8463	1.1468	0.5816	0.9433	0.5894	0.6075	0.5625	0.5283	0.5441	0.2729
<b>CTMAPE</b>																								
REC-MLP	0.8336	1.0702	1.9696	2.7360	3.6581	4.4215	5.4213	9.7053	10.706	16.323	16.075	15.280	14.954	14.394	12.190	11.160	8.5652	8.0832	1.5718	0.3986	1.3511	1.9119	0.9058	0.4476
REC-LIN	0.8463	1.0846	2.0064	2.7868	3.7226	4.5003	5.5128	9.8054	10.818	16.438	16.202	15.419	15.104	14.555	12.366	11.348	8.7685	8.2974	1.8116	0.1438	1.0838	1.6333	0.6204	0.1543
DIR-MLP	0.8463	0.5876	0.5939	<b>0.1051</b>	<b>0.4751</b>	1.3343	2.0014	<b>0.9418</b>	<b>0.4405</b>	<b>5.3605</b>	<b>4.0935</b>	2.5883	2.1263	1.9202	<b>0.2789</b>	<b>0.4045</b>	<b>0.9007</b>	<b>0.2636</b>	4.9739	5.2634	4.5541	3.9566	1.4631	0.2641
DIR-LIN	0.8352	0.5988	1.9749	2.2965	2.5237	3.9540	2.9653	5.0902	7.5562	10.646	6.4008	<b>2.4626</b>	<b>1.6701</b>	<b>0.2609</b>	2.3848	2.7758	5.6625	5.6442	12.510	11.355	9.9884	3.9956	1.9979	<b>0.1480</b>
ARIMA	<b>0.0391</b>	<b>0.1813</b>	1.1436	1.6572	2.3807	3.3017	4.4707	8.6704	9.5532	15.363	15.247	14.357	13.924	13.458	11.356	10.247	7.5278	7.1279	<b>0.6722</b>	1.3755	2.4303	2.9197	1.7932	1.3750
RTY-BST-L2 BI	0.8463	0.3216	0.1632	0.1075	0.9314	0.6506	0.5566	3.6149	3.4579	8.6522	6.5753	3.1347	2.0136	1.0048	1.4185	1.6783	2.8812	1.2180	5.9042	5.4860	4.4869	3.8090	1.8363	0.4107
RTY-BST-L1 BI	1.1098	0.3166	<b>0.1632</b>	0.3279	0.6989	0.5783	0.8293	4.1585	4.2968	9.8019	10.572	6.5841	3.2944	2.1408	1.0597	1.0658	2.7534	0.3378	4.2007	4.2356	3.4868	2.4851	<b>0.5235</b>	0.7969
RTY-BST-L2 UN	0.8853	0.5356	0.9975	1.1686	1.1444	<b>0.4323</b>	<b>0.2964</b>	2.4706	1.9096	7.8639	8.1905	8.1863	9.0263	8.2614	5.4587	6.4742	4.5639	5.9401	0.8416	<b>0.0318</b>	<b>0.3294</b>	<b>0.7534</b>	0.6641	2.4547
RTY-BST-L1 UN	0.3779	1.9976	0.2518	0.5650	0.7773	0.7248	0.5010	2.6299	2.7891	7.2230	3.9530	8.4074	8.3664	10.139	7.0138	9.1768	4.7434	7.4373	4.4794	4.5246	3.9028	4.0537	2.0844	
<b>MASE</b>																								
REC-MLP	0.5699	0.7317	1.3462	1.8697	2.4995	3.0205	3.7029	6.6279	7.3107	11.143	10.972	10.428	10.204	9.8209	8.3162	7.6126	5.8414	5.5119	1.0717	0.2717	0.9209	1.3031	0.6172	0.3049
REC-LIN	0.5787	0.7417	1.3719	1.9054	2.5453	3.0768	3.7691	6.7037	7.3959	11.238	11.076	10.540	10.325	9.9495	8.4528	7.7570	5.9933	5.6712	1.2382	0.0982	0.7407	1.1162	0.4240	0.1054
DIR-MLP	0.5711	0.4075	1.3499	1.5623	1.7043	2.6880	1.9741	3.3071	4.9835	6.8063	3.9175	<b>1.4598</b>	<b>0.9857</b>	<b>0.1528</b>	1.3953	1.6366	3.3418	3.3486	7.4623	6.9796	6.2738	2.6866	1.3469	<b>0.1011</b>
DIR-LIN	0.5787	0.3998	0.4003	<b>0.0699</b>	<b>0.3113</b>	0.8598	1.2675	<b>0.5862</b>	<b>0.2697</b>	<b>3.2357</b>	<b>2.4451</b>	1.5363	1.2608	1.1435	<b>0.1675</b>	<b>0.2481</b>	<b>0.5566</b>	<b>0.1656</b>	3.1797	3.4223	3.0091	2.4372	0.9916	0.1802
ARIMA	<b>0.0265</b>	<b>0.1229</b>	0.7751	1.1200	1.6054	2.2294	3.0232	5.8540	6.4397	10.369	10.306	9.6934	9.3877	9.0833	7.6742	6.9184	5.0763	4.8105	<b>0.4541</b>	0.9286	1.6391	1.9704	1.2113	0.9283
RTY-BST-L2 BI	0.5787	0.5042	0.7095	0.6833	0.6189	0.4276	0.3616	2.3127	2.1838	5.4107	4.0318	1.8711	1.1926	0.5928	0.8378	1.0002	1.7463	0.7542	3.7413	3.5595	2.9666	2.5485	1.2399	0.2799
RTY-BST-L1 BI	0.7609	0.2168	<b>0.1092</b>	0.2187	0.4633	0.3798	0.5402	2.6755	2.7373	6.2079	6.7712	1.9770	1.2777	0.6281	0.6390	1.6710	0.2124	2.7054	2.7812	2.3276	1.6842	<b>0.3581</b>	0.5411	
RTY-BST-L2 UN	0.6056	0.3765	0.6751	0.7859	0.7621	<b>0.2835</b>	<b>0.1909</b>	1.5620	1.1869	4.8757	5.1105	5.1553	5.7579	5.2598	3.4586	4.1947	2.9820	3.9583	0.5695	<b>0.0217</b>	<b>0.2268</b>	<b>0.5194</b>	0.4597	1.7224
RTY-BST-L1 UN	0.2552	1.3788	0.1691	0.3734	0.5077	0.4698	0.3220	1.6655	1.7493	4.4474	3.6261	5.3073	5.2986	6.5903	4.5183	6.1226	3.1052	5.0361	3.1471	3.2436	3.0035	2.8209	2.9052	1.4570

Tabela B.3: Medidas de precisão globais das previsões para as diferentes estratégias no consumo de energia elétrica em Cabinda.

Estratégias	Métricas		
	RMSE	MAPE (%)	MASE
REC-MLP	1.1357	6.8390	4.6675
REC-LIN	1.1502	6.8763	4.7006
DIR-MLP	0.6612	4.4041	2.7677
DIR-LIN	0.8176	5.3263	3.6165
ARIMA	1.0388	6.2740	4.2353
RTY-BST-L2 BI	<b>0.4047</b>	<b>2.6405</b>	<b>1.6731</b>
RTY-BST-L1 BI	0.4681	2.7426	1.7483
RTY-BST-L2 UNI	0.5425	3.2874	2.1125
RTY-BST-L1 UNI	0.6420	4.2619	2.8133

Tabela B.4: Valor real observado juntamente com os valores previstos dos retornos da NYSE ao longo dos horizontes de previsões cinco à frente  $H = 5$ .

Estratégias	Horizonte de Previsão ( $h$ )				
	1	2	3	4	5
<b>REAL</b>	<b>0.011772</b>	<b>0.011380</b>	<b>0.005903</b>	<b>-0.00152</b>	<b>0.001376</b>
REC-MLP	0.000462	0.000462	0.000462	0.000462	0.000462
REC-LIN	0.001209	0.000692	0.000546	0.000491	0.000474
DIR-MLP	0.000468	0.000464	0.000464	0.000467	0.000463
DIR-LIN	0.001209	0.000419	0.000343	0.000278	0.000488
GARCH	0.000902	0.000748	0.000732	0.000730	0.000730
RTY-BST-L2 BI	0.000760	0.000463	0.000309	0.000542	0.000623
RTY-BST-L1 BI	-0.23355	0.133451	0.134517	-0.16377	-0.17766
RTY-BST-L2 UN	0.000959	0.000334	0.000516	0.000392	0.000444
RTY-BST-L1 UN	-0.21299	0.207659	-0.21164	-0.17033	-0.18064

Tabela B.5: Medidas de precisão ao longo dos horizontes de previsões cinco passos à frente  $H = 5$  dos retornos da NYSE.

Estratégias	Horizonte de Previsão ( $h$ )				
	1	2	3	4	5
<b>RMSE</b>					
REC-MLP	0.011310	0.010918	0.005440	0.001987	0.000913
REC-LIN	<b>0.010563</b>	0.010688	0.005356	0.002016	0.000901
DIR-MLP	0.011304	0.010915	0.005438	0.001992	0.000913
DIR-LIN	<b>0.010563</b>	0.010960	0.005559	<b>0.001803</b>	0.000888
GARCH	0.010870	<b>0.010632</b>	<b>0.005170</b>	0.002256	<b>0.000645</b>
RTY-BST-L2 BI	0.011012	0.010916	0.005593	0.002067	0.000752
RTY-BST-L1 BI	0.245328	0.122070	0.128614	0.162251	0.179045
RTY-BST-L2 UN	0.010813	0.011046	0.005386	0.001917	0.000931
RTY-BST-L1 UN	0.224770	0.196278	0.217546	0.168810	0.182025
<b>MASE</b>					
REC-MLP	1.27340	1.22923	0.61253	0.22380	0.10290
REC-LIN	<b>1.18928</b>	1.20334	0.60310	0.22706	0.10154
DIR-MLP	1.27276	1.22895	0.61228	0.22436	0.10283
DIR-LIN	<b>1.18928</b>	1.23405	0.62593	<b>0.20308</b>	0.10000
GARCH	1.22387	<b>1.19708</b>	<b>0.58215</b>	0.25401	<b>0.07271</b>
RTY-BST-L2 BI	1.23981	1.22908	0.62978	0.23281	0.08477
RTY-BST-L1 BI	27.62070	13.74349	14.48029	18.26732	20.15813
RTY-BST-L2 UN	1.21745	1.24368	0.60641	0.21590	0.10492
RTY-BST-L1 UN	25.30612	22.09832	24.49285	19.00581	20.49363

Tabela B.6: Medidas de precisão das diferentes estratégias e modelos de previsão para  $H = 5$  dos retornos da NYSE.

Estratégias	Métricas	
	RMSE	MASE
REC-MLP	0.007503	0.68837
REC-LIN	<b>0.007202</b>	<b>0.66486</b>
DIR-MLP	0.007501	0.68823
DIR-LIN	0.007303	0.67047
GARCH	0.007258	0.66596
RTY-BST-L2 BI	0.007437	0.68325
RTY-BST-L1 BI	0.173212	18.85398
RTY-BST-L2 UN	0.007382	0.67767
RTY-BST-L1 UN	0.198999	22.27935

Tabela B.7: Valor real observado juntamente com os valores previstos dos índices de produção mensal do Sistema de Reserva Federal com  $H = 12$ .

Estratégias	Horizonte de Previsão ( $h$ )											
	1	2	3	4	5	6	7	8	9	10	11	12
<b>REAL</b>	<b>134.8</b>	<b>139.6</b>	<b>141.4</b>	<b>144.2</b>	<b>144.2</b>	<b>148.8</b>	<b>141.9</b>	<b>146.9</b>	<b>152</b>	<b>152.6</b>	<b>149.7</b>	<b>145</b>
REC-MLP	135.11	137.52	134.90	135.61	134.49	134.67	134.13	134.15	133.87	133.84	133.68	133.64
REC-LIN	135.22	135.37	134.67	134.73	137.51	140.17	141.07	140.72	141.24	141.65	141.01	140.01
DIR-MLP	129.63	133.11	135.08	133.45	133.77	130.54	133.40	134.46	134.60	134.05	129.44	135.30
DIR-LIN	135.22	139.08	140.10	141.22	139.48	138.34	138.31	139.84	143.01	143.92	141.69	138.10
ARIMA	135.41	139.61	140.44	140.76	141.74	145.61	138.84	143.62	147.51	147.26	144.43	139.65
RTY-BST-L2 BI	134.30	137.64	138.14	136.30	132.45	134.86	134.26	132.55	133.20	135.81	133.18	131.52
RTY-BST-L1 BI	135.17	136.35	137.80	136.68	137.57	140.08	139.63	138.96	139.82	139.54	138.83	139.00
RTY-BST-L2 UN	134.40	135.38	135.56	134.11	129.57	128.99	130.80	132.68	135.93	140.51	138.75	134.09
RTY-BST-L1 UN	135.25	134.39	137.00	135.03	134.78	137.40	139.29	140.04	141.51	142.18	141.31	140.67

Tabela B.8: Medidas de precisão ao longo dos horizontes de previsões doze passos à frente  $H = 12$  do índice de produção mensal do Sistema de Reserva Federal.

Estratégias	Horizonte de Previsão ( $h$ )											
	1	2	3	4	5	6	7	8	9	10	11	12
<b>RMSE</b>												
REC-MLP	<b>0.317</b>	2.077	6.490	8.587	9.705	14.123	7.760	12.745	18.127	18.758	16.015	11.357
REC-LIN	0.421	4.226	6.729	9.462	6.681	8.625	<b>0.821</b>	6.175	10.750	10.943	8.686	4.988
DIR-MLP	5.160	6.481	6.312	10.746	10.421	18.252	8.492	12.434	17.398	18.549	20.252	9.692
DIR-LIN	0.421	0.513	1.291	<b>2.977</b>	4.717	10.45	3.582	7.051	8.989	8.679	8.005	6.899
ARIMA	0.611	<b>0.016</b>	<b>0.958</b>	3.437	<b>2.455</b>	<b>3.188</b>	3.053	<b>3.272</b>	<b>4.482</b>	<b>5.331</b>	<b>5.268</b>	5.343
RTY-BST-L2 BI	0.492	1.957	3.253	7.893	11.740	13.936	7.630	14.342	18.793	16.787	16.518	13.470
RTY-BST-L1 BI	0.373	3.241	3.593	7.514	6.622	8.714	2.268	7.939	12.171	13.053	10.860	5.995
RTY-BST-L2 UN	0.397	4.212	5.837	10.087	14.622	19.800	11.091	14.210	16.067	12.088	10.944	10.906
RTY-BST-L1 UN	0.452	5.208	4.399	9.167	9.416	11.395	2.600	6.858	10.480	10.410	8.383	<b>4.324</b>
<b>MAPE</b>												
REC-MLP	<b>0.235</b>	1.510	4.810	6.332	7.216	10.486	5.785	9.500	13.539	14.015	11.980	8.498
REC-LIN	0.311	3.121	4.996	7.022	4.858	6.153	<b>0.582</b>	4.388	7.610	7.725	6.160	3.562
DIR-MLP	3.980	4.869	4.672	8.052	7.790	13.981	6.366	9.247	12.926	13.837	15.645	7.163
DIR-LIN	0.311	0.369	0.921	<b>2.108</b>	3.382	7.556	2.590	5.042	6.286	6.031	5.649	4.996
ARIMA	0.451	<b>0.011</b>	<b>0.682</b>	2.442	<b>1.732</b>	<b>2.189</b>	2.199	<b>2.278</b>	<b>3.038</b>	<b>3.619</b>	<b>3.648</b>	3.825
RTY-BST-L2 BI	0.366	1.421	2.354	5.790	8.863	10.333	5.683	10.819	14.108	12.361	12.402	10.241
RTY-BST-L1 BI	0.276	2.377	2.607	5.497	4.813	6.220	1.624	5.713	8.704	9.354	7.822	4.313
RTY-BST-L1 UN	0.295	3.111	4.306	7.521	11.284	15.349	8.479	10.709	11.819	8.603	7.887	8.833
RTY-BST-L2 UN	0.334	3.875	3.211	6.789	6.986	8.293	1.866	4.897	7.405	7.321	5.932	<b>3.074</b>
<b>MASE</b>												
REC-MLP	<b>0.161</b>	1.056	3.299	4.365	4.933	7.179	3.945	6.479	9.214	9.536	8.141	5.773
REC-LIN	0.214	2.148	3.420	4.810	3.396	4.384	<b>0.417</b>	3.139	5.464	5.563	4.415	2.535
DIR-MLP	2.623	3.295	3.208	5.462	5.298	9.279	4.317	6.321	8.844	9.429	10.29	4.927
DIR-LIN	0.214	0.260	0.656	<b>1.513</b>	2.398	5.314	1.821	3.584	4.570	4.412	4.069	3.507
ARIMA	0.310	<b>0.008</b>	<b>0.487</b>	1.747	<b>1.248</b>	<b>1.620</b>	1.552	<b>1.663</b>	<b>2.278</b>	<b>2.710</b>	<b>2.678</b>	2.716
RTY-BST-L2 BI	0.250	0.994	1.653	4.012	5.968	7.084	3.879	7.290	9.553	8.534	8.397	6.848
RTY-BST-L1 BI	0.189	1.647	1.826	3.819	3.366	4.430	1.153	4.036	6.187	6.635	5.521	3.048
RTY-BST-L2 UN	0.202	2.141	2.967	5.128	7.433	10.065	5.638	7.224	8.167	6.145	5.563	5.544
RTY-BST-L1 UN	0.230	2.647	2.236	4.660	4.786	5.792	1.321	3.486	5.327	5.292	4.261	<b>2.198</b>

Tabela B.9: Medidas de precisão das diferentes estratégias e modelos de previsão para  $H = 12$  do índice de produção mensal.

Estratégias	Métricas		
	RMSE	MAPE	MASE
REC-MLP	11.911	7.8260	5.3405
REC-LIN	7.3393	4.7079	3.3260
DIR-MLP	13.049	9.0444	6.1086
DIR-LIN	6.2778	3.7704	2.6936
ARIMA	<b>3.5841</b>	<b>2.1766</b>	<b>1.5851</b>
RTY-BST-L2 BI	12.118	7.8957	5.3723
RTY-BST-L1 BI	7.8552	4.9438	3.4886
RTY-BST-L2 UN	11.998	8.1252	5.5185
RTY-BST-L1 UN	7.6932	4.9989	3.5202